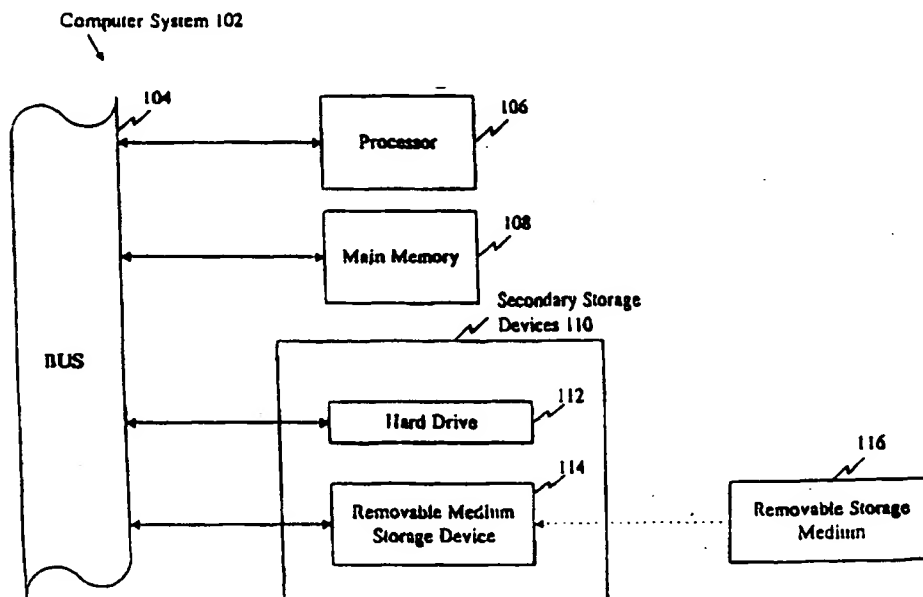




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : C12N 15/31, C07K 14/315, 16/12, C12Q 1/68		A2	(11) International Publication Number: WO 98/18931
(21) International Application Number: PCT/US97/19588		(43) International Publication Date: 7 May 1998 (07.05.98)	
(22) International Filing Date: 30 October 1997 (30.10.97)		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	
(30) Priority Data: 60/029,960 31 October 1996 (31.10.96) US		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).		<p>Published</p> <p>Without international search report and to be republished upon receipt of that report.</p>	
(72) Inventors; and			
(75) Inventors/Applicants (for US only): KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).			

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES

## (57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.





## *Streptococcus pneumoniae* Polynucleotides and Sequences

### FIELD OF THE INVENTION

5           The present invention relates to the field of molecular biology. In particular, it relates to, among other things, nucleotide sequences of *Streptococcus pneumoniae*, contigs, ORFs, fragments, probes, primers and related polynucleotides thereof, peptides and polypeptides encoded by the sequences, and uses of the polynucleotides and sequences thereof, such as in fermentation,  
10       polypeptide production, assays and pharmaceutical development, among others.

### BACKGROUND OF THE INVENTION

*Streptococcus pneumoniae* has been one of the most extensively studied  
15       microorganisms since its first isolation in 1881. It was the object of many investigations that led to important scientific discoveries. In 1928, Griffith observed that when heat-killed encapsulated pneumococci and live strains constitutively lacking any capsule were concomitantly injected into mice, the nonencapsulated could be converted into encapsulated pneumococci with the same  
20       capsular type as the heat-killed strain. Years later, the nature of this "transforming principle," or carrier of genetic information, was shown to be DNA. (Avery, O.T., et al., *J. Exp. Med.*, 79:137-157 (1944)).

          In spite of the vast number of publications on *S. pneumoniae* many questions about its virulence are still unanswered, and this pathogen remains a  
25       major causative agent of serious human disease, especially community-acquired pneumonia. (Johnston, R.B., et al., *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991)). In addition, in developing countries, the pneumococcus is responsible for the death of a large number of children under the age of 5 years from pneumococcal pneumonia. The incidence of pneumococcal disease is highest in infants under 2  
30       years of age and in people over 60 years of age. Pneumococci are the second most frequent cause (after *Haemophilus influenzae* type b) of bacterial meningitis and otitis media in children. With the recent introduction of conjugate vaccines for *H. influenzae* type b, pneumococcal meningitis is likely to become increasingly prominent. *S. pneumoniae* is the most important etiologic agent of community-

acquired pneumonia in adults and is the second most common cause of bacterial meningitis behind *Neisseria meningitidis*.

The antibiotic generally prescribed to treat *S. pneumoniae* is benzylpenicillin, although resistance to this and to other antibiotics is found occasionally. Pneumococcal resistance to penicillin results from mutations in its penicillin-binding proteins. In uncomplicated pneumococcal pneumonia caused by a sensitive strain, treatment with penicillin is usually successful unless started too late. Erythromycin or clindamycin can be used to treat pneumonia in patients hypersensitive to penicillin, but resistant strains to these drugs exist. Broad spectrum antibiotics (e.g., the tetracyclines) may also be effective, although tetracycline-resistant strains are not rare. In spite of the availability of antibiotics, the mortality of pneumococcal bacteremia in the last four decades has remained stable between 25 and 29%. (Gillespie, S.H., et al., *J. Med. Microbiol.* 28:237-248 (1989).

*S. pneumoniae* is carried in the upper respiratory tract by many healthy individuals. It has been suggested that attachment of pneumococci is mediated by a disaccharide receptor on fibronectin, present on human pharyngeal epithelial cells. (Anderson, B.J., et al., *J. Immunol.* 142:2464-2468 (1989). The mechanisms by which pneumococci translocate from the nasopharynx to the lung, thereby causing pneumonia, or migrate to the blood, giving rise to bacteremia or septicemia, are poorly understood. (Johnston, R.B., et al., *Rev. Infect. Dis.* 13(Suppl. 6):S509-517 (1991).

Various proteins have been suggested to be involved in the pathogenicity of *S. pneumoniae*, however, only a few of them have actually been confirmed as virulence factors. Pneumococci produce an IgA1 protease that might interfere with host defense at mucosal surfaces. (Kornfield, S.J., et al., *Rev. Inf. Dis.* 3:521-534 (1981). *S. pneumoniae* also produces neuraminidase, an enzyme that may facilitate attachment to epithelial cells by cleaving sialic acid from the host glycolipids and gangliosides. Partially purified neuraminidase was observed to induce meningitis-like symptoms in mice; however, the reliability of this finding has been questioned because the neuraminidase preparations used were probably contaminated with cell wall products. Other pneumococcal proteins besides neuraminidase are involved in the adhesion of pneumococci to epithelial and endothelial cells. These pneumococcal proteins have as yet not been identified. Recently, Cundell et al., reported that peptide permeases can modulate

pneumococcal adherence to epithelial and endothelial cells. It was, however, unclear whether these permeases function directly as adhesions or whether they enhance adherence by modulating the expression of pneumococcal adhesions. (DeVelasco, E.A., *et al.*, *Micro. Rev.* 59:591-603 (1995). A better understanding of the virulence factors determining its pathogenicity will need to be developed to cope with the devastating effects of pneumococcal disease in humans.

Ironically, despite the prominent role of *S. pneumoniae* in the discovery of DNA, little is known about the molecular genetics of the organism. The *S. pneumoniae* genome consists of one circular, covalently closed, double-stranded DNA and a collection of so-called variable accessory elements, such as prophages, plasmids, transposons and the like. Most physical characteristics and almost all of the genes of *S. pneumoniae* are unknown. Among the few that have been identified, most have not been physically mapped or characterized in detail. Only a few genes of this organism have been sequenced. (See, for instance current versions of GENBANK and other nucleic acid databases, and references that relate to the genome of *S. pneumoniae* such as those set out elsewhere herein.)

It is clear that the etiology of diseases mediated or exacerbated by *S. pneumoniae*, infection involves the programmed expression of *S. pneumoniae* genes, and that characterizing the genes and their patterns of expression would add dramatically to our understanding of the organism and its host interactions. Knowledge of *S. pneumoniae* genes and genomic organization would improve our understanding of disease etiology and lead to improved and new ways of preventing, ameliorating, arresting and reversing diseases. Moreover, characterized genes and genomic fragments of *S. pneumoniae* would provide reagents for, among other things, detecting, characterizing and controlling *S. pneumoniae* infections. There is a need to characterize the genome of *S. pneumoniae* and for polynucleotides of this organism.

## SUMMARY OF THE INVENTION

5 The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome. The primary nucleotide sequences which were generated are provided in SEQ ID NOS:1-391.

10 The present invention provides the nucleotide sequence of several hundred contigs of the *Streptococcus pneumoniae* genome, which are listed in tables below and set out in the Sequence Listing submitted herewith, and representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan. In one embodiment, the present invention is provided as contiguous strings of primary sequence information corresponding to the nucleotide sequences depicted in SEQ ID NOS:1-391.

15 The present invention further provides nucleotide sequences which are at least 95% identical to the nucleotide sequences of SEQ ID NOS:1-391.

20 The nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NOS:1-391 may be provided in a variety of mediums to facilitate its use. In one application of this embodiment, the sequences of the present invention are recorded on computer readable media. Such media includes, but is not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

25 The present invention further provides systems, particularly computer-based systems which contain the sequence information herein described stored in a data storage means. Such systems are designed to identify commercially important fragments of the *Streptococcus pneumoniae* genome.

30 Another embodiment of the present invention is directed to fragments of the *Streptococcus pneumoniae* genome having particular structural or functional attributes. Such fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to, fragments which encode peptides, hereinafter referred to as open reading frames or ORFs, fragments which modulate the expression of an operably linked ORF, hereinafter referred to as expression modulating fragments or EMFs, and fragments which can be used to diagnose the

35

presence of *Streptococcus pneumoniae* in a sample, hereinafter referred to as diagnostic fragments or DFs.

Each of the ORFs in fragments of the *Streptococcus pneumoniae* genome disclosed in Tables 1-3, and the EMFs found 5' to the ORFs, can be used in numerous ways as polynucleotide reagents. For instance, the sequences can be used as diagnostic probes or amplification primers for detecting or determining the presence of a specific microbe in a sample, to selectively control gene expression in a host and in the production of polypeptides, such as polypeptides encoded by ORFs of the present invention, particular those polypeptides that have a pharmacological activity.

The present invention further includes recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genome of the present invention. The recombinant constructs of the present invention comprise vectors, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* has been inserted.

The present invention further provides host cells containing any of the isolated fragments of the *Streptococcus pneumoniae* genome of the present invention. The host cells can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic cell, such as a yeast cell, or a procaryotic cell such as a bacterial cell.

The present invention is further directed to isolated polypeptides and proteins encoded by ORFs of the present invention. A variety of methods, well known to those of skill in the art, routinely may be utilized to obtain any of the polypeptides and proteins of the present invention. For instance, polypeptides and proteins of the present invention having relatively short, simple amino acid sequences readily can be synthesized using commercially available automated peptide synthesizers. Polypeptides and proteins of the present invention also may be purified from bacterial cells which naturally produce the protein. Yet another alternative is to purify polypeptide and proteins of the present invention from cells which have been altered to express them.

The invention further provides methods of obtaining homologs of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. Specifically, by using the nucleotide and amino acid sequences disclosed herein as

a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

The invention further provides antibodies which selectively bind polypeptides and proteins of the present invention. Such antibodies include both  
5 monoclonal and polyclonal antibodies.

The invention further provides hybridomas which produce the above-described antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

The present invention further provides methods of identifying test samples  
10 derived from cells which express one of the ORFs of the present invention, or a homolog thereof. Such methods comprise incubating a test sample with one or more of the antibodies of the present invention, or one or more of the DFs of the present invention, under conditions which allow a skilled artisan to determine if the sample contains the ORF or product produced therefrom.

15 In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the above-described assays.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the antibodies, or one of the DFs of the present invention; and  
20 (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of bound antibodies or hybridized DFs.

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents capable of binding to  
25 a polypeptide or protein encoded by one of the ORFs of the present invention. Specifically, such agents include, as further described below, antibodies, peptides, carbohydrates, pharmaceutical agents and the like. Such methods comprise steps of: (a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention; and (b) determining whether the agent binds to said protein.

30 The present genomic sequences of *Streptococcus pneumoniae* will be of great value to all laboratories working with this organism and for a variety of commercial purposes. Many fragments of the *Streptococcus pneumoniae* genome will be immediately identified by similarity searches against GenBank or protein databases and will be of immediate value to *Streptococcus pneumoniae* researchers

and for immediate commercial value for the production of proteins or to control gene expression.

The methodology and technology for elucidating extensive genomic sequences of bacterial and other genomes has and will greatly enhance the ability to analyze and understand chromosomal organization. In particular, sequenced contigs and genomes will provide the models for developing tools for the analysis of chromosome structure and function, including the ability to identify genes within large segments of genomic DNA, the structure, position, and spacing of regulatory elements, the identification of genes with potential industrial applications, and the ability to do comparative genomic and molecular phylogeny.

### **DESCRIPTION OF THE FIGURES**

**FIGURE 1** is a block diagram of a computer system (102) that can be used to implement computer-based systems of present invention.

**FIGURE 2** is a schematic diagram depicting the data flow and computer programs used to collect, assemble, edit and annotate the contigs of the *Streptococcus pneumoniae* genome of the present invention. Both Macintosh and Unix platforms are used to handle the AB 373 and 377 sequence data files, largely as described in Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, 585, IEEE Computer Society Press, Washington D.C. (1993). Factura (AB) is a Macintosh program designed for automatic vector sequence removal and end-trimming of sequence files. The program Loadis runs on a Macintosh platform and parses the feature data extracted from the sequence files by Factura to the Unix based *Streptococcus pneumoniae* relational database. Assembly of contigs (and whole genome sequences) is accomplished by retrieving a specific set of sequence files and their associated features using Extrseq, a Unix utility for retrieving sequences from an SQL database. The resulting sequence file is processed by seq\_filter to trim portions of the sequences with more than 2% ambiguous nucleotides. The sequence files were assembled using TIGR Assembler, an assembly engine designed at The Institute for Genomic Research ( TIGR ) for rapid and accurate assembly of thousands of sequence fragments. The collection of contigs generated by the assembly step is loaded into the database with the lassie program. Identification of open reading

frames (ORFs) is accomplished by processing contigs with zorf or GenMark. The ORFs are searched against *S. pneumoniae* sequences from GenBank and against all protein sequences using the BLASTN and BLASTP programs, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)). Results of the ORF determination and similarity searching steps were loaded into the database. As described below, some results of the determination and the searches are set out in Tables 1-3.

### DETAILED DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

10

The present invention is based on the sequencing of fragments of the *Streptococcus pneumoniae* genome and analysis of the sequences. The primary nucleotide sequences generated by sequencing the fragments are provided in SEQ ID NOS:1-391. (As used herein, the "primary sequence" refers to the nucleotide sequence represented by the IUPAC nomenclature system.)

15

In addition to the aforementioned *Streptococcus pneumoniae* polynucleotide and polynucleotide sequences, the present invention provides the nucleotide sequences of SEQ ID NOS:1-391, or representative fragments thereof, in a form which can be readily used, analyzed, and interpreted by a skilled artisan.

20

As used herein, a "representative fragment of the nucleotide sequence depicted in SEQ ID NOS:1-391" refers to any portion of the SEQ ID NOS:1-391 which is not presently represented within a publicly available database. Preferred representative fragments of the present invention are *Streptococcus pneumoniae* open reading frames ( ORFs ), expression modulating fragment ( EMFs ) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in sample ( DFs ). A non-limiting identification of preferred representative fragments is provided in Tables 1-3. As discussed in detail below, the information provided in SEQ ID NOS:1-391 and in Tables 1-3 together with routine cloning, synthesis, sequencing and assay methods will enable those skilled in the art to clone and sequence all "representative fragments" of interest, including open reading frames encoding a large variety of *Streptococcus pneumoniae* proteins.

25

30

While the presently disclosed sequences of SEQ ID NOS:1-391 are highly accurate, sequencing techniques are not perfect and, in relatively rare instances, further investigation of a fragment or sequence of the invention may reveal a

35



nucleotide sequence error present in a nucleotide sequence disclosed in SEQ ID NOS:1-391. However, once the present invention is made available (*i.e.*, once the information in SEQ ID NOS:1-391 and Tables 1-3 has been made available), resolving a rare sequencing error in SEQ ID NOS:1-391 will be well within the skill of the art. The present disclosure makes available sufficient sequence information to allow any of the described contigs or portions thereof to be obtained readily by straightforward application of routine techniques. Further sequencing of such polynucleotide may proceed in like manner using manual and automated sequencing methods which are employed ubiquitous in the art. Nucleotide sequence editing software is publicly available. For example, Applied Biosystem's (AB) AutoAssembler can be used as an aid during visual inspection of nucleotide sequences. By employing such routine techniques potential errors readily may be identified and the correct sequence then may be ascertained by targeting further sequencing effort, also of a routine nature, to the region containing the potential error.

Even if all of the very rare sequencing errors in SEQ ID NOS:1-391 were corrected, the resulting nucleotide sequences would still be at least 95% identical, nearly all would be at least 99% identical, and the great majority would be at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391.

As discussed elsewhere herein, polynucleotides of the present invention readily may be obtained by routine application of well known and standard procedures for cloning and sequencing DNA. Detailed methods for obtaining libraries and for sequencing are provided below, for instance. A wide variety of *Streptococcus pneumoniae* strains that can be used to prepare *S. pneumoniae* genomic DNA for cloning and for obtaining polynucleotides of the present invention are available to the public from recognized depository institutions, such as the American Type Culture Collection (ATCC). While the present invention is enabled by the sequences and other information herein disclosed, the *S. pneumoniae* strain that provided the DNA of the present Sequence Listing, Strain 7/87 14.8.91, has been deposited in the ATCC, as a convenience to those of skill in the art. As a further convenience, a library of *S. pneumoniae* genomic DNA, derived from the same strain, also has been deposited in the ATCC. The *S. pneumoniae* strain was deposited on October 10, 1996, and was given Deposit No. 55840, and the cDNA library was deposited on October 11, 1996 and was given Deposit No. 97755. The genomic fragments in the library are 15 to 20 kb

fragments generated by partial Sau3A1 digestion and they are inserted into the BamHI site in the well-known lambda-derived vector lambda DASH II (Stratagene, La Jolla, CA). The provision of the deposits is not a waiver of any rights of the inventors or their assignees in the present subject matter.

5       The nucleotide sequences of the genomes from different strains of *Streptococcus pneumoniae* differ somewhat. However, the nucleotide sequences of the genomes of all *Streptococcus pneumoniae* strains will be at least 95% identical, in corresponding part, to the nucleotide sequences provided in SEQ ID NOS:1-391. Nearly all will be at least 99% identical and the great majority will be  
10   99.9% identical.

Thus, the present invention further provides nucleotide sequences which are at least 95%, preferably 99% and most preferably 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391, in a form which can be readily used, analyzed and interpreted by the skilled artisan.

15       Methods for determining whether a nucleotide sequence is at least 95%, at least 99% or at least 99.9% identical to the nucleotide sequences of SEQ ID NOS:1-391 are routine and readily available to the skilled artisan. For example, the well known fasta algorithm described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988) can be used to generate the percent identity of nucleotide  
20   sequences. The BLASTN program also can be used to generate an identity score of polynucleotides compared to one another.

### COMPUTER RELATED EMBODIMENTS

The nucleotide sequences provided in SEQ ID NOS:1-391, a representative  
25   fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide sequence of SEQ ID NOS:1-391 may be "provided" in a variety of mediums to facilitate use thereof. As used herein, provided refers to a manufacture, other than an isolated nucleic acid molecule, which contains a nucleotide sequence of the present invention; i.e.,  
30   a nucleotide sequence provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a polynucleotide of SEQ ID NOS:1-391. Such a manufacture provides a large portion of the *Streptococcus pneumoniae* genome and parts thereof (e.g., a *Streptococcus pneumoniae* open reading frame  
35   (ORF)) in a form which allows a skilled artisan to examine the manufacture using

means not directly applicable to examining the *Streptococcus pneumoniae* genome or a subset thereof as it exists in nature or in purified form.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium which can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories, such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention. Likewise, it will be clear to those of skill how additional computer readable media that may be developed also can be used to create analogous manufactures having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data-processor structuring formats (e.g., text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. Thus, by providing in computer readable form the nucleotide sequences of SEQ ID NOS:1-

391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most preferably at least 99.9% identical to a sequence of SEQ ID NOS:1-391 the present invention enables the skilled artisan routinely to access the provided sequence information for a wide variety of purposes.

5       The examples which follow demonstrate how software which implements the BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990)) and BLAZE (Brutlag *et al.*, *Comp. Chem.* 17:203-207 (1993)) search algorithms on a Sybase system was used to identify open reading frames (ORFs) within the *Streptococcus pneumoniae* genome which contain homology to ORFs or proteins from both  
10 *Streptococcus pneumoniae* and from other organisms. Among the ORFs discussed herein are protein encoding fragments of the *Streptococcus pneumoniae* genome useful in producing commercially important proteins, such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

15       The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify, among other things, commercially important fragments of the *Streptococcus pneumoniae* genome.

20       As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention.

25       As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means.

30       As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

35       As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage

means. Search means are used to identify fragments or regions of the present genomic sequences which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the *Streptococcus pneumoniae* genomic sequences possessing varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments of the

*Streptococcus pneumoniae* genome. In the present examples, implementing software which implement the BLAST and BLAZE algorithms, described in Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990), is used to identify open reading frames within the *Streptococcus pneumoniae* genome. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Of course, suitable proprietary systems that may be known to those of skill also may be employed in this regard.

Figure 1 provides a block diagram of a computer system illustrative of embodiments of this aspect of present invention. The computer system 102 includes a processor 106 connected to a bus 104. Also connected to the bus 104 are a main memory 108 (preferably implemented as random access memory, RAM) and a variety of secondary storage devices 110, such as a hard drive 112 and a removable medium storage device 114. The removable medium storage device 114 may represent, for example, a floppy disk drive, a CD-ROM drive, a magnetic tape drive, *etc.* A removable storage medium 116 (such as a floppy disk, a compact disk, a magnetic tape, *etc.*) containing control logic and/or data recorded therein may be inserted into the removable medium storage device 114. The computer system 102 includes appropriate software for reading the control logic and/or the data from the removable medium storage device 114, once it is inserted into the removable medium storage device 114.

A nucleotide sequence of the present invention may be stored in a well known manner in the main memory 108, any of the secondary storage devices 110, and/or a removable storage medium 116. During execution, software for accessing and processing the genomic sequence (such as search tools, comparing tools, *etc.*) reside in main memory 108, in accordance with the requirements and operating parameters of the operating system, the hardware system and the software program or programs.

## BIOCHEMICAL EMBODIMENTS

Other embodiments of the present invention are directed to isolated fragments of the *Streptococcus pneumoniae* genome. The fragments of the *Streptococcus pneumoniae* genome of the present invention include, but are not limited to fragments which encode peptides and polypeptides, hereinafter open reading frames (ORFs), fragments which modulate the expression of an operably linked ORF, hereinafter expression modulating fragments (EMFs) and fragments which can be used to diagnose the presence of *Streptococcus pneumoniae* in a sample, hereinafter diagnostic fragments (DFs).

As used herein, an "isolated nucleic acid molecule" or an "isolated fragment of the *Streptococcus pneumoniae* genome" refers to a nucleic acid molecule possessing a specific nucleotide sequence which has been subjected to purification means to reduce, from the composition, the number of compounds which are normally associated with the composition. Particularly, the term refers to the nucleic acid molecules having the sequences set out in SEQ ID NOS:1-391, to representative fragments thereof as described above, to polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence thereto, also as set out above.

A variety of purification means can be used to generate the isolated fragments of the present invention. These include, but are not limited to methods which separate constituents of a solution based on charge, solubility, or size.

In one embodiment, *Streptococcus pneumoniae* DNA can be enzymatically sheared to produce fragments of 15-20 kb in length. These fragments can then be used to generate a *Streptococcus pneumoniae* library by inserting them into lambda clones as described in the Examples below. Primers flanking, for example, an ORF, such as those enumerated in Tables 1-3 can then be generated using nucleotide sequence information provided in SEQ ID NOS:1-391. Well known and routine techniques of PCR cloning then can be used to isolate the ORF from the lambda DNA library or *Streptococcus pneumoniae* genomic DNA. Thus, given the availability of SEQ ID NOS:1-391, the information in Tables 1, 2 and 3, and the information that may be obtained readily by analysis of the sequences of SEQ ID NOS:1-391 using methods set out above, those of skill will be enabled by the present disclosure to isolate any ORF-containing or other nucleic acid fragment of the present invention.

The isolated nucleic acid molecules of the present invention include, but are not limited to single stranded and double stranded DNA, and single stranded RNA.

As used herein, an "open reading frame," ORF, means a series of triplets coding for amino acids without any termination codons and is a sequence  
5 translatable into protein.

Tables 1, 2, and 3 list ORFs in the *Streptococcus pneumoniae* genomic contigs of the present invention that were identified as putative coding regions by the GeneMark software using organism-specific second-order Markov probability transition matrices. It will be appreciated that other criteria can be used, in  
10 accordance with well known analytical methods, such as those discussed herein, to generate more inclusive, more restrictive, or more selective lists.

Table 1 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that over a continuous region of at least 50 bases are 95% or more identical (by BLAST analysis) to a nucleotide sequence available through  
15 GenBank in October, 1997.

Table 2 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that are not in Table 1 and match, with a BLASTP probability score of 0.01 or less, a polypeptide sequence available through GenBank in October, 1997.

20 Table 3 sets out ORFs in the *Streptococcus pneumoniae* contigs of the present invention that do not match significantly, by BLASTP analysis, a polypeptide sequence available through GenBank in October, 1997.

In each table, the first and second columns identify the ORF by, respectively, contig number and ORF number within the contig; the third column  
25 indicates the first nucleotide of the ORF (actually the first nucleotide of the stop codon immediately preceeding the ORF), counting from the 5' end of the contig strand; and the fourth column, "stop (nt)" indicates the last nucleotide of the stop codon defining the 3' end of the ORF.

In Tables 1 and 2, column five, lists the Reference for the closest  
30 matching sequence available through GenBank. These reference numbers are the databases entry numbers commonly used by those of skill in the art, who will be familiar with their denominators. Descriptions of the nomenclature are available from the National Center for Biotechnology Information. Column six in Tables 1 and 2 provides the gene name of the matching sequence; column seven provides  
35 the BLAST identity score and column eight the BLAST similarity score from the



comparison of the ORF and the homologous gene; and column nine indicates the length in nucleotides of the highest scoring segment pair identified by the BLAST identity analysis.

Each ORF described in the tables is defined by "start (nt)" (5') and "stop (nt)" (3') nucleotide position numbers. These position numbers refer to the boundaries of each ORF and provide orientation with respect to whether the forward or reverse strand is the coding strand and which reading frame the coding sequence is contained. The "start" position is the first nucleotide of the triplet encoding a stop codon just 5' to the ORF and the "stop" position is the last nucleotide of the triplet encoding the next in-frame stop codon (i.e., the stop codon at the 3' end of the ORF). Those of ordinary skill in the art appreciate that preferred fragments within each ORF described in the table include fragments of each ORF which include the entire sequence from the delineated "start" and "stop" positions excepting the first and last three nucleotides since these encode stop codons. Thus, polynucleotides set out as ORFs in the tables but lacking the three (3) 5' nucleotides and the three (3) 3' nucleotides are encompassed by the present invention. Those of skill also appreciate that particularly preferred are fragments within each ORF that are polynucleotide fragments comprising polypeptide coding sequence. As defined herein, "coding sequence" includes the fragment within an ORF beginning at the first in-frame ATG (triplet encoding methionine) and ending with the last nucleotide prior to the triplet encoding the 3' stop codon. Preferred are fragments comprising the entire coding sequence and fragments comprising the entire coding sequence, excepting the coding sequence for the N-terminal methionine. Those of skill appreciate that the N-terminal methionine is often removed during post-translational processing and that polynucleotides lacking the ATG can be used to facilitate production of N-terminal fusion proteins which may be beneficial in the production or use of genetically engineered proteins. Of course, due to the degeneracy of the genetic code many polynucleotides can encode a given polypeptide. Thus, the invention further includes polynucleotides comprising a nucleotide sequence encoding a polypeptide sequence itself encoded by the coding sequence within an ORF described in Tables 1-3 herein. Further, polynucleotides at least 95%, preferably at least 99% and especially preferably at least 99.9% identical in sequence to the foregoing polynucleotides, are contemplated by the present invention.

Polypeptides encoded by polynucleotides described above and elsewhere herein are also provided by the present invention as are polypeptide comprising a an amino acid sequence at least about 95%, preferably at least 97% and even more preferably 99% identical to the amino acid sequence of a polypeptide encoded by an ORF shown in Tables 1-3. These polypeptides may or may not comprise an N-terminal methionine.

The concepts of percent identity and percent similarity of two polypeptide sequences is well understood in the art. For example, two polypeptides 10 amino acids in length which differ at three amino acid positions (*e.g.*, at positions 1, 3 and 5) are said to have a percent identity of 70%. However, the same two polypeptides would be deemed to have a percent similarity of 80% if, for example at position 5, the amino acids moieties, although not identical, were "similar" (*i.e.*, possessed similar biochemical characteristics). Many programs for analysis of nucleotide or amino acid sequence similarity, such as fasta and BLAST specifically list percent identity of a matching region as an output parameter. Thus, for instance, Tables 1 and 2 herein enumerate the percent identity of the highest scoring segment pair in each ORF and its listed relative. Further details concerning the algorithms and criteria used for homology searches are provided below and are described in the pertinent literature highlighted by the citations provided below.

It will be appreciated that other criteria can be used to generate more inclusive and more exclusive listings of the types set out in the tables. As those of skill will appreciate, narrow and broad searches both are useful. Thus, a skilled artisan can readily identify ORFs in contigs of the *Streptococcus pneumoniae* genome other than those listed in Tables 1-3, such as ORFs which are overlapping or encoded by the opposite strand of an identified ORF in addition to those ascertainable using the computer-based systems of the present invention.

As used herein, an "expression modulating fragment," EMF, means a series of nucleotide molecules which modulates the expression of an operably linked ORF or EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

EMF sequences can be identified within the contigs of the *Streptococcus pneumoniae* genome by their proximity to the ORFs provided in Tables 1-3. An intergenic segment, or a fragment of the intergenic segment, from about 10 to 200 nucleotides in length, taken from any one of the ORFs of Tables 1-3 will modulate the expression of an operably linked ORF in a fashion similar to that found with the naturally linked ORF sequence. As used herein, an "intergenic segment" refers to fragments of the *Streptococcus pneumoniae* genome which are between two ORF(s) herein described. EMFs also can be identified using known EMFs as a target sequence or target motif in the computer-based systems of the present invention. Further, the two methods can be combined and used together.

The presence and activity of an EMF can be confirmed using an EMF trap vector. An EMF trap vector contains a cloning site linked to a marker sequence. A marker sequence encodes an identifiable phenotype, such as antibiotic resistance or a complementing nutrition auxotrophic factor, which can be identified or assayed when the EMF trap vector is placed within an appropriate host under appropriate conditions. As described above, a EMF will modulate the expression of an operably linked marker sequence. A more detailed discussion of various marker sequences is provided below. A sequence which is suspected as being an EMF is cloned in all three reading frames in one or more restriction sites upstream from the marker sequence in the EMF trap vector. The vector is then transformed into an appropriate host using known procedures and the phenotype of the transformed host is examined under appropriate conditions. As described above, an EMF will modulate the expression of an operably linked marker sequence.

As used herein, a "diagnostic fragment," DF, means a series of nucleotide molecules which selectively hybridize to *Streptococcus pneumoniae* sequences. DFs can be readily identified by identifying unique sequences within contigs of the *Streptococcus pneumoniae* genome, such as by using well-known computer analysis software, and by generating and testing probes or amplification primers

consisting of the DF sequence in an appropriate diagnostic format which determines amplification or hybridization selectivity.

The sequences falling within the scope of the present invention are not limited to the specific sequences herein described, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequences provided in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95%, preferably at least 99% and most at least preferably 99.9% identical to SEQ ID NOS:1-391, with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another which encodes the same amino acid is expressly contemplated. Any specific sequence disclosed herein can be readily screened for errors by resequencing a particular fragment, such as an ORF, in both directions (*i.e.*, sequence both strands). Alternatively, error screening can be performed by sequencing corresponding polynucleotides of *Streptococcus pneumoniae* origin isolated by using part or all of the fragments in question as a probe or primer.

Preferred DFs of the present invention comprise at least about 17, preferably at least about 20, and more preferably at least about 50 contiguous nucleotides within an ORF set out in Tables 1-3. Most highly preferred DFs specifically hybridize to a polynucleotide containing the sequence of the ORF from which they are derived. Specific hybridization occurs even under stringent conditions defined elsewhere herein.

Each of the ORFs of the *Streptococcus pneumoniae* genome disclosed in Tables 1, 2 and 3, and the EMFs found 5' to the ORFs, can be used as polynucleotide reagents in numerous ways. For example, the sequences can be used as diagnostic probes or diagnostic amplification primers to detect the presence of a specific microbe in a sample, particularly *Streptococcus pneumoniae*. Especially preferred in this regard are ORFs such as those of Table 3, which do not match previously characterized sequences from other organisms and thus are most likely to be highly selective for *Streptococcus pneumoniae*. Also particularly preferred are ORFs that can be used to distinguish between strains of *Streptococcus pneumoniae*, particularly those that distinguish medically important strain, such as drug-resistant strains.

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Information from the sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription, for triple-helix formation, or to the mRNA itself, for antisense inhibition. Both techniques have been demonstrated to be effective in model systems, and the requisite techniques are well known and involve routine procedures. Triple helix techniques are discussed in, for example, Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991). Antisense techniques in general are discussed in, for instance, Okano, *J. Neurochem.* 56:560 (1991) and *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)).

The present invention further provides recombinant constructs comprising one or more fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention. Certain preferred recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a fragment of the *Streptococcus pneumoniae* genome has been inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences, including for example, a promoter, operably linked to the ORF. For vectors comprising the EMFs of the present invention, the vector may further comprise a marker sequence or heterologous ORF operably linked to the EMF.

Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Useful bacterial vectors include phagescript, PsiX174, pBluescript SK, pBS KS, pNH8a, pNH16a, pNH18a, pNH46a (available from Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (available from Pharmacia). Useful eukaryotic vectors include pWLneo, pSV2cat, pOG44, pXT1, pSG

(available from Stratagene) pSVK3, pBPV, pMSG, pSVL (available from Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

5 Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein- I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

10 The present invention further provides host cells containing any one of the isolated fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, wherein the fragment has been introduced into the host cell using known methods. The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or  
15 a procaryotic cell, such as a bacterial cell.

A polynucleotide of the present invention, such as a recombinant construct comprising an ORF of the present invention, may be introduced into the host by a variety of well established techniques that are standard in the art, such as calcium phosphate transfection, DEAE, dextran mediated transfection and electroporation,  
20 which are described in, for instance, Davis, L. *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986).

A host cell containing one of the fragments of the *Streptococcus pneumoniae* genomic fragments and contigs of the present invention, can be used in conventional manners to produce the gene product encoded by the isolated  
25 fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF. The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ  
30 from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the Genetic Code, encode an identical polypeptide sequence.

Preferred nucleic acid fragments of the present invention are the ORFs and subfragments thereof depicted in Tables 2 and 3 which encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. This is particularly useful in producing small peptides and fragments of larger polypeptides. —Such short fragments as may be obtained most readily by synthesis are useful, for example, in generating antibodies against the native polypeptide, as discussed further below.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily employ well-known methods for isolating polypeptides and proteins to isolate and purify polypeptides or proteins of the present invention produced naturally by a bacterial strain, or by other methods. Methods for isolation and purification that can be employed in this regard include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography.

The polypeptides and proteins of the present invention also can be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. Those skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, CV-1 cell, COS cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level.

"Recombinant," as used herein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (*e.g.*, yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern different from that expressed in mammalian cells.

"Nucleotide sequence" refers to a heteropolymer of deoxyribonucleotides. Generally, DNA segments encoding the polypeptides and proteins provided by this invention are assembled from fragments of the *Streptococcus pneumoniae* genome and short oligonucleotide linkers, or from a series of oligonucleotides, to provide a synthetic gene which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon.

Recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. The expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic regulatory elements necessary for gene expression in the host, including elements required to initiate and maintain transcription at a level sufficient for suitable expression of the desired polypeptide, including, for example, promoters and, where necessary, an enhancer and a polyadenylation signal; (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate signals to initiate translation at the beginning of the desired coding region and terminate translation at its end. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

"Recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extra chromosomally. The cells can be prokaryotic or eukaryotic. Recombinant expression systems as defined herein will express



heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference in its entirety.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), alpha-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and, when desirable, provide amplification within the host.

Suitable prokaryotic hosts for transformation include strains of *E. coli*, *B. subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas* and *Streptomyces*. Others may, also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication

derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (available from Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (available from Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter, where it is inducible, is derepressed or induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period to provide for expression of the induced gene product. Thereafter cells are typically harvested, generally by centrifugation, disrupted to release expressed protein, generally by physical or chemical means, and the resulting crude extract is retained for further purification.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described in Gluzman, *Cell* 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines.

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Recombinant polypeptides and proteins produced in bacterial culture is usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The present invention further includes isolated polypeptides, proteins and nucleic acid molecules which are substantially equivalent to those herein described. As used herein, substantially equivalent can refer both to nucleic acid and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between reference and subject sequences. For purposes of the present invention, sequences having equivalent biological activity, and equivalent expression characteristics are considered substantially equivalent. For purposes of determining equivalence, truncation of the mature sequence should be disregarded.

The invention further provides methods of obtaining homologs from other strains of *Streptococcus pneumoniae*, of the fragments of the *Streptococcus pneumoniae* genome of the present invention and homologs of the proteins encoded by the ORFs of the present invention. As used herein, a sequence or protein of *Streptococcus pneumoniae* is defined as a homolog of a fragment of the *Streptococcus pneumoniae* fragments or contigs or a protein encoded by one of the ORFs of the present invention, if it shares significant homology to one of the fragments of the *Streptococcus pneumoniae* genome of the present invention or a protein encoded by one of the ORFs of the present invention. Specifically, by using the sequence disclosed herein as a probe or as primers, and techniques such as PCR cloning and colony/plaque hybridization, one skilled in the art can obtain homologs.

As used herein, two nucleic acid molecules or proteins are said to "share significant homology" if the two contain regions which possess greater than 85% sequence (amino acid or nucleic acid) homology. Preferred homologs in this regard are those with more than 90% homology. Especially preferred are those with 93% or more homology. Among especially preferred homologs those with 95% or more homology are particularly preferred. Very particularly preferred among these are those with 97% and even more particularly preferred among these are homologs with 99% or more homology. The most preferred homologs among these are those with 99.9% homology or more. It will be understood that, among measures of homology, identity is particularly preferred in this regard.

Region specific primers or probes derived from the nucleotide sequence provided in SEQ ID NOS:1-391 or from a nucleotide sequence at least 95%, particularly at least 99%, especially at least 99.5% identical to a sequence of SEQ

ID NOS:1-391 can be used to prime DNA synthesis and PCR amplification, as well as to identify colonies containing cloned DNA encoding a homolog. Methods suitable to this aspect of the present invention are well known and have been described in great detail in many publications such as, for example, Innis *et al.*,  
5 *PCR Protocols*, Academic Press, San Diego, CA (1990)).

When using primers derived from SEQ ID NOS:1-391 or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, annealing at 50-60°C in 6X SSPC and 50% formamide, and washing at 50-  
10 65°C in 0.5X SSPC) only sequences which are greater than 75% homologous to the primer will be amplified. By employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences which are greater than 40-50% homologous to the primer will also be amplified.

15 When using DNA probes derived from SEQ ID NOS:1-391, or from a nucleotide sequence having an aforementioned identity to a sequence of SEQ ID NOS:1-391, for colony/plaque hybridization, one skilled in the art will recognize that by employing high stringency conditions (*e.g.*, hybridizing at 50- 65°C in 5X SSPC and 50% formamide, and washing at 50- 65°C in 0.5X SSPC), sequences  
20 having regions which are greater than 90% homologous to the probe can be obtained, and that by employing lower stringency conditions (*e.g.*, hybridizing at 35-37°C in 5X SSPC and 40-45% formamide, and washing at 42°C in 0.5X SSPC), sequences having regions which are greater than 35-45% homologous to the probe will be obtained.

25 Any organism can be used as the source for homologs of the present invention so long as the organism naturally expresses such a protein or contains genes encoding the same. The most preferred organism for isolating homologs are bacteria which are closely related to *Streptococcus pneumoniae*.

## 30 ILLUSTRATIVE USES OF COMPOSITIONS OF THE INVENTION

Each ORF provided in Tables 1 and 2 is identified with a function by homology to a known gene or polypeptide. As a result, one skilled in the art can use the polypeptides of the present invention for commercial, therapeutic and  
35 industrial purposes consistent with the type of putative identification of the

polypeptide. Such identifications permit one skilled in the art to use the *Streptococcus pneumoniae* ORFs in a manner similar to the known type of sequences for which the identification is made; for example, to ferment a particular sugar source or to produce a particular metabolite. A variety of reviews illustrative of this aspect of the invention are available, including the following reviews on the industrial use of enzymes, for example, BIOCHEMICAL ENGINEERING AND BIOTECHNOLOGY HANDBOOK, 2nd Ed., MacMillan Publications, Ltd., NY (1991) and BIOCATALYSTS IN ORGANIC SYNTHESSES, Tramper *et al.*, Eds., Elsevier Science Publishers, Amsterdam, The Netherlands (1985). A variety of exemplary uses that illustrate this and similar aspects of the present invention are discussed below.

### 1. Biosynthetic Enzymes

Open reading frames encoding proteins involved in mediating the catalytic reactions involved in intermediary and macromolecular metabolism, the biosynthesis of small molecules, cellular processes and other functions includes enzymes involved in the degradation of the intermediary products of metabolism, enzymes involved in central intermediary metabolism, enzymes involved in respiration, both aerobic and anaerobic, enzymes involved in fermentation, enzymes involved in ATP proton motor force conversion, enzymes involved in broad regulatory function, enzymes involved in amino acid synthesis, enzymes involved in nucleotide synthesis, enzymes involved in cofactor and vitamin synthesis, can be used for industrial biosynthesis.

The various metabolic pathways present in *Streptococcus pneumoniae* can be identified based on absolute nutritional requirements as well as by examining the various enzymes identified in Table 1-3 and SEQ ID NOS:1-391.

Of particular interest are polypeptides involved in the degradation of intermediary metabolites as well as non-macromolecular metabolism. Such enzymes include amylases, glucose oxidases, and catalase.

Proteolytic enzymes are another class of commercially important enzymes. Proteolytic enzymes find use in a number of industrial processes including the processing of flax and other vegetable fibers, in the extraction, clarification and depectinization of fruit juices, in the extraction of vegetables' oil and in the maceration of fruits and vegetables to give unicellular fruits. A detailed review of the proteolytic enzymes used in the food industry is provided in Rombouts *et al.*,

*Symbiosis* 21:79 (1986) and Voragen *et al.* in *Biocatalysts In Agricultural Biotechnology*, Whitaker *et al.*, Eds., *American Chemical Society Symposium Series* 389:93 (1989).

The metabolism of sugars is an important aspect of the primary metabolism of *Streptococcus pneumoniae*. Enzymes involved in the degradation of sugars, such as, particularly, glucose, galactose, fructose and xylose, can be used in industrial fermentation. Some of the important sugar transforming enzymes, from a commercial viewpoint, include sugar isomerases such as glucose isomerase. Other metabolic enzymes have found commercial use such as glucose oxidases which produces ketogulonic acid (KGA). KGA is an intermediate in the commercial production of ascorbic acid using the Reichstein's procedure, as described in Krueger *et al.*, *Biotechnology* 6(A), Rhine *et al.*, Eds., Verlag Press, Weinheim, Germany (1984).

Glucose oxidase (GOD) is commercially available and has been used in purified form as well as in an immobilized form for the deoxygenation of beer. See, for instance, Hartmeir *et al.*, *Biotechnology Letters* 1:21 (1979). The most important application of GOD is the industrial scale fermentation of gluconic acid. Market for gluconic acids which are used in the detergent, textile, leather, photographic, pharmaceutical, food, feed and concrete industry, as described, for example, in Bigelis *et al.*, beginning on page 357 in *GENE MANIPULATIONS AND FUNGI*; Benett *et al.*, Eds., Academic Press, New York (1985). In addition to industrial applications, GOD has found applications in medicine for quantitative determination of glucose in body fluids recently in biotechnology for analyzing syrups from starch and cellulose hydrosylates. This application is described in Owusu *et al.*, *Biochem. et Biophysica. Acta.* 872:83 (1986), for instance.

The main sweetener used in the world today is sugar which comes from sugar beets and sugar cane. In the field of industrial enzymes, the glucose isomerase process shows the largest expansion in the market today. Initially, soluble enzymes were used and later immobilized enzymes were developed (Krueger *et al.*, *Biotechnology, The Textbook of Industrial Microbiology*, Sinauer Associated Incorporated, Sunderland, Massachusetts (1990)). Today, the use of glucose- produced high fructose syrups is by far the largest industrial business using immobilized enzymes. A review of the industrial use of these enzymes is provided by Jorgensen, *Starch* 40:307 (1988).

Proteinases, such as alkaline serine proteinases, are used as detergent additives and thus represent one of the largest volumes of microbial enzymes used in the industrial sector. Because of their industrial importance, there is a large body of published and unpublished information regarding the use of these enzymes in industrial processes. (See Faultman *et al.*, *Acid Proteases Structure Function and Biology*, Tang, J., ed., Plenum Press, New York (1977) and Godfrey *et al.*, *Industrial Enzymes*, MacMillan Publishers, Surrey, UK (1983) and Hepner *et al.*, *Report Industrial Enzymes by 1990*, Hel Hepner & Associates, London (1986)).

Another class of commercially usable proteins of the present invention are the microbial lipases, described by, for instance, Macrae *et al.*, *Philosophical Transactions of the Chiral Society of London* 310:227 (1985) and Poserke, *Journal of the American Oil Chemist Society* 61:1758 (1984). A major use of lipases is in the fat and oil industry for the production of neutral glycerides using lipase catalyzed inter-esterification of readily available triglycerides. Application of lipases include the use as a detergent additive to facilitate the removal of fats from fabrics in the course of the washing procedures.

The use of enzymes, and in particular microbial enzymes, as catalyst for key steps in the synthesis of complex organic molecules is gaining popularity at a great rate. One area of great interest is the preparation of chiral intermediates. Preparation of chiral intermediates is of interest to a wide range of synthetic chemists particularly those scientists involved with the preparation of new pharmaceuticals, agrochemicals, fragrances and flavors. (See Davies *et al.*, *Recent Advances in the Generation of Chiral Intermediates Using Enzymes*, CRC Press, Boca Raton, Florida (1990)). The following reactions catalyzed by enzymes are of interest to organic chemists: hydrolysis of carboxylic acid esters, phosphate esters, amides and nitriles, esterification reactions, trans-esterification reactions, synthesis of amides, reduction of alkanones and oxoalkanates, oxidation of alcohols to carbonyl compounds, oxidation of sulfides to sulfoxides, and carbon bond forming reactions such as the aldol reaction.

When considering the use of an enzyme encoded by one of the ORFs of the present invention for biotransformation and organic synthesis it is sometimes necessary to consider the respective advantages and disadvantages of using a microorganism as opposed to an isolated enzyme. Pros and cons of using a whole cell system on the one hand or an isolated partially purified enzyme on the other

hand, has been described in detail by Bud *et al.*, Chemistry in Britain (1987), p. 127.

Amino transferases, enzymes involved in the biosynthesis and metabolism of amino acids, are useful in the catalytic production of amino acids. The advantages of using microbial based enzyme systems is that the amino transferase enzymes catalyze the stereo- selective synthesis of only L-amino acids and generally possess uniformly high catalytic rates. A description of the use of amino transferases for amino acid production is provided by Roselle-David, *Methods of Enzymology* 136:479 (1987).

Another category of useful proteins encoded by the ORFs of the present invention include enzymes involved in nucleic acid synthesis, repair, and recombination.

## 2. Generation of Antibodies

As described here, the proteins of the present invention, as well as homologs thereof, can be used in a variety of procedures and methods known in the art which are currently applied to other proteins. The proteins of the present invention can further be used to generate an antibody which selectively binds the protein. Such antibodies can be either monoclonal or polyclonal antibodies, as well fragments of these antibodies, and humanized forms.

The invention further provides antibodies which selectively bind to one of the proteins of the present invention and hybridomas which produce these antibodies. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody.

In general, techniques for preparing polyclonal and monoclonal antibodies as well as hybridomas capable of producing the desired antibody are well known in the art (Campbell, A. M., *Monoclonal Antibody Technology: Laboratory Techniques In Biochemistry And Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984); St. Groth *et al.*, *J. Immunol. Methods* 35: 1-21 (1980), Kohler and Milstein, *Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* 4:72 (1983), pgs. 77-96 of Cole *et al.*, in *Monoclonal Antibodies And Cancer Therapy*, Alan R. Liss, Inc. (1985)). Any animal (mouse, rabbit, etc.) which is known to produce antibodies can be immunized with the pseudogene polypeptide. Methods for immunization are well known in the art. Such methods



include subcutaneous or interperitoneal injection of the polypeptide. One skilled in the art will recognize that the amount of the protein encoded by the ORF of the present invention used for immunization will vary based on the animal which is immunized, the antigenicity of the peptide and the site of injection.

5       The protein which is used as an immunogen may be modified or administered in an adjuvant in order to increase the protein's antigenicity. Methods of increasing the antigenicity of a protein are well known in the art and include, but are not limited to coupling the antigen with a heterologous protein (such as globulin or galactosidase) or through the inclusion of an adjuvant during immunization.

10       For monoclonal antibodies, spleen cells from the immunized animals are removed, fused with myeloma cells, such as SP2/0-Ag14 myeloma cells, and allowed to become monoclonal antibody producing hybridoma cells.

Any one of a number of methods well known in the art can be used to identify the hybridoma cell which produces an antibody with the desired characteristics. These include screening the hybridomas with an ELISA assay, western blot analysis, or radioimmunoassay (Lutz *et al.*, *Exp. Cell Res.* 175:109-124 (1988)).

Hybridomas secreting the desired antibodies are cloned and the class and subclass is determined using procedures known in the art (Campbell, A. M., 20 *Monoclonal Antibody Technology: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1984)).

Techniques described for the production of single chain antibodies (U. S. Patent 4,946,778) can be adapted to produce single chain antibodies to proteins of 25 the present invention.

For polyclonal antibodies, antibody containing antisera is isolated from the immunized animal and is screened for the presence of antibodies with the desired specificity using one of the above-described procedures.

The present invention further provides the above- described antibodies in 30 detectably labelled form. Antibodies can be detectably labelled through the use of radioisotopes, affinity labels (such as biotin, avidin, *etc.*), enzymatic labels (such as horseradish peroxidase, alkaline phosphatase, *etc.*) fluorescent labels (such as FITC or rhodamine, *etc.*), paramagnetic atoms, *etc.* Procedures for accomplishing such labeling are well-known in the art, for example see Sternberger *et al.*, *J. Histochem. Cytochem.* 18:315 (1970); Bayer, E. A. *et al.*, *Meth. Enzym.* 62:308 35

(1979); Engval, E. *et al.*, *Immunol.* 109:129 (1972); Goding, J. W., *J. Immunol. Meth.* 13:215 (1976)).

The labeled antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays to identify cells or tissues in which a fragment of the  
5 *Streptococcus pneumoniae* genome is expressed.

The present invention further provides the above-described antibodies immobilized on a solid support. Examples of such solid supports include plastics such as polycarbonate, complex carbohydrates such as agarose and sepharose, acrylic resins and such as polyacrylamide and latex beads. Techniques for  
10 coupling antibodies to such solid supports are well known in the art (Weir, D. M. *et al.*, "Handbook of Experimental Immunology" 4th Ed., Blackwell Scientific Publications, Oxford, England, Chapter 10 (1986); Jacoby, W. D. *et al.*, *Meth. Enzym.* 34 Academic Press, N. Y. (1974)). The immobilized antibodies of the present invention can be used for *in vitro*, *in vivo*, and in situ assays as well as for  
15 immunoaffinity purification of the proteins of the present invention.

### 3. Diagnostic Assays and Kits

The present invention further provides methods to identify the expression of one of the ORFs of the present invention, or homolog thereof, in a test sample,  
20 using one of the DFs or antibodies of the present invention.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the DFs of the present invention and assaying for binding of the DFs or antibodies to components within the test sample.

Conditions for incubating a DF or antibody with a test sample vary.  
25 Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the DF or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the DFs or antibodies of the present invention. Examples of such assays  
30 can be found in Chard, T., *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and*

*Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the DFs or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound DF or antibody.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, *etc.*), and containers which contain the reagents used to detect the bound antibody or DF.

Types of detection reagents include labelled nucleic acid probes, labelled secondary antibodies, or in the alternative, if the primary antibody is labelled, the enzymatic, or antibody binding reagents which are capable of reacting with the labelled antibody. One skilled in the art will readily recognize that the disclosed DFs and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

#### 4. Screening Assay for Binding Agents

Using the isolated proteins of the present invention, the present invention further provides methods of obtaining and identifying agents which bind to a protein encoded by one of the ORFs of the present invention or to one of the fragments and the *Streptococcus pneumoniae* fragment and contigs herein described.

In general, such methods comprise steps of:

(a) contacting an agent with an isolated protein encoded by one of the ORFs of the present invention, or an isolated fragment of the *Streptococcus pneumoniae* genome; and

(b) determining whether the agent binds to said protein or said fragment.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention.

Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like capable of binding to a specific peptide sequence in order to generate rationally designed antipeptide peptides, for example see Hurby *et al.*, "Application of Synthetic Peptides: Antisense Peptides," in *Synthetic Peptides, A User's Guide*, W. H. Freeman, NY (1992), pp. 289-307, and Kaspczak *et al.*, *Biochemistry* 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control.

One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

5 Agents suitable for use in these methods usually contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee *et al.*, *Nucl. Acids Res.* 6:3073 (1979); Cooney *et al.*, *Science* 241:456 (1988); and Dervan *et al.*, *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991);  
10 *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix- formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the  
15 sequences of the present invention can be used to design antisense and triple helix-forming oligonucleotides, and other DNA binding agents.

## 5. Pharmaceutical Compositions and Vaccines

The present invention further provides pharmaceutical agents which can be  
20 used to modulate the growth or pathogenicity of *Streptococcus pneumoniae*, or another related organism, *in vivo* or *in vitro*. As used herein, a "pharmaceutical agent" is defined as a composition of matter which can be formulated using known techniques to provide a pharmaceutical compositions. As used herein, the "pharmaceutical agents of the present invention" refers the pharmaceutical agents  
25 which are derived from the proteins encoded by the ORFs of the present invention or are agents which are identified using the herein described assays.

As used herein, a pharmaceutical agent is said to "modulate the growth pathogenicity of *Streptococcus pneumoniae* or a related organism, *in vivo* or *in vitro*," when the agent reduces the rate of growth, rate of division, or viability of  
30 the organism in question. The pharmaceutical agents of the present invention can modulate the growth or pathogenicity of an organism in many fashions, although an understanding of the underlying mechanism of action is not needed to practice the use of the pharmaceutical agents of the present invention. Some agents will modulate the growth by binding to an important protein thus blocking the biological  
35 activity of the protein, while other agents may bind to a component of the outer

surface of the organism blocking attachment or rendering the organism more prone to act the bodies nature immune system. Alternatively, the agent may comprise a protein encoded by one of the ORFs of the present invention and serve as a vaccine. The development and use of a vaccine based on outer membrane components are well known in the art.

As used herein, a "related organism" is a broad term which refers to any organism whose growth can be modulated by one of the pharmaceutical agents of the present invention. In general, such an organism will contain a homolog of the protein which is the target of the pharmaceutical agent or the protein used as a vaccine. As such, related organisms do not need to be bacterial but may be fungal or viral pathogens.

The pharmaceutical agents and compositions of the present invention may be administered in a convenient manner, such as by the oral, topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, they are administered in an amount of at least about 1 mg/kg body weight and in most cases they will be administered in an amount not in excess of about 1 g/kg body weight per day. In most cases, the dosage is from about 0.1 mg/kg to about 10 g/kg body weight daily, taking into account the routes of administration, symptoms, *etc.*

The agents of the present invention can be used in native form or can be modified to form a chemical derivative. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, *etc.* The moieties may alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, *etc.* Moieties capable of mediating such effects are disclosed in, among other sources, REMINGTON'S PHARMACEUTICAL SCIENCES (1980) cited elsewhere herein.

For example, such moieties may change an immunological character of the functional derivative, such as affinity for a given antibody. Such changes in immunomodulation activity are measured by the appropriate assay, such as a competitive type immunoassay. Modifications of such protein properties as redox or thermal stability, biological half-life, hydrophobicity, susceptibility to proteolytic degradation or the tendency to aggregate with carriers or into multimers also may

be effected in this way and can be assayed by methods well known to the skilled artisan.

The therapeutic effects of the agents of the present invention may be obtained by providing the agent to a patient by any suitable means (*e.g.*, inhalation, intravenously, intramuscularly, subcutaneously, enterally, or parenterally). It is preferred to administer the agent of the present invention so as to achieve an effective concentration within the blood or tissue in which the growth of the organism is to be controlled. To achieve an effective blood concentration, the preferred method is to administer the agent by injection. The administration may be by continuous infusion, or by single or multiple injections.

In providing a patient with one of the agents of the present invention, the dosage of the administered agent will vary depending upon such factors as the patient's age, weight, height, sex, general medical condition, previous medical history, *etc.* In general, it is desirable to provide the recipient with a dosage of agent which is in the range of from about 1 pg/kg to 10 mg/kg (body weight of patient), although a lower or higher dosage may be administered. The therapeutically effective dose can be lowered by using combinations of the agents of the present invention or another agent.

As used herein, two or more compounds or agents are said to be administered "in combination" with each other when either (1) the physiological effects of each compound, or (2) the serum concentrations of each compound can be measured at the same time. The composition of the present invention can be administered concurrently with, prior to, or following the administration of the other agent.

The agents of the present invention are intended to be provided to recipient subjects in an amount sufficient to decrease the rate of growth (as defined above) of the target organism.

The administration of the agent(s) of the invention may be for either a "prophylactic" or "therapeutic" purpose. When provided prophylactically, the agent(s) are provided in advance of any symptoms indicative of the organisms growth. The prophylactic administration of the agent(s) serves to prevent, attenuate, or decrease the rate of onset of any subsequent infection. When provided therapeutically, the agent(s) are provided at (or shortly after) the onset of an indication of infection. The therapeutic administration of the compound(s)

serves to attenuate the pathological symptoms of the infection and to increase the rate of recovery.

The agents of the present invention are administered to a subject, such as a mammal, or a patient, in a pharmaceutically acceptable form and in a therapeutically effective concentration. A composition is said to be "pharmacologically acceptable" if its administration can be tolerated by a recipient patient. Such an agent is said to be administered in a "therapeutically effective amount" if the amount administered is physiologically significant. An agent is physiologically significant if its presence results in a detectable change in the physiology of a recipient patient.

The agents of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby these materials, or their functional derivatives, are combined in a mixture with a pharmaceutically acceptable carrier vehicle. Suitable vehicles and their formulation, inclusive of other human proteins, *e.g.*, human serum albumin, are described, for example, in REMINGTON'S PHARMACEUTICAL SCIENCES, 16<sup>th</sup> Ed., Osol, A., Ed., Mack Publishing, Easton PA (1980). In order to form a pharmaceutically acceptable composition suitable for effective administration, such compositions will contain an effective amount of one or more of the agents of the present invention, together with a suitable amount of carrier vehicle.

Additional pharmaceutical methods may be employed to control the duration of action. Control release preparations may be achieved through the use of polymers to complex or absorb one or more of the agents of the present invention. The controlled delivery may be effectuated by a variety of well known techniques, including formulation with macromolecules such as, for example, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine, sulfate, adjusting the concentration of the macromolecules and the agent in the formulation, and by appropriate use of methods of incorporation, which can be manipulated to effectuate a desired time course of release. Another possible method to control the duration of action by controlled release preparations is to incorporate agents of the present invention into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization with, for example, hydroxymethylcellulose or gelatine-



microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions. Such techniques are disclosed in REMINGTON'S PHARMACEUTICAL SCIENCES  
5 (1980).

The invention further provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or  
10 sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

In addition, the agents of the present invention may be employed in conjunction with other therapeutic compounds.

## 15 6. Shot-Gun Approach to Megabase DNA Sequencing

The present invention further demonstrates that a large sequence can be sequenced using a random shotgun approach. This procedure, described in detail in the examples that follow, has eliminated the up front cost of isolating and ordering overlapping or contiguous subclones prior to the start of the sequencing  
20 protocols.

Certain aspects of the present invention are described in greater detail in the examples that follow. The examples are provided by way of illustration. Other aspects and embodiments of the present invention are contemplated by the inventors, as will be clear to those of skill in the art from reading the present  
25 disclosure.

## ILLUSTRATIVE EXAMPLES

### LIBRARIES AND SEQUENCING

#### 30 1. Shotgun Sequencing Probability Analysis

The overall strategy for a shotgun approach to whole genome sequencing follows from the Lander and Waterman (Landerman and Waterman, *Genomics* 2:231 (1988)) application of the equation for the Poisson distribution. According to this treatment, the probability,  $P$ , that any given base in a sequence of size  $L$ , in  
35 nucleotides, is not sequenced after a certain amount,  $n$ , in nucleotides, of random

sequence has been determined can be calculated by the equation  $P = e^{-m}$ , where  $m$  is  $L/n$ , the fold coverage. For instance, for a genome of 2.8 Mb,  $m=1$  when 2.8 Mb of sequence has been randomly generated (1X coverage). At that point,  $P = e^{-1} = 0.37$ . The probability that any given base has not been sequenced is the same as the probability that any region of the whole sequence  $L$  has not been determined and, therefore, is equivalent to the fraction of the whole sequence that has yet to be determined. Thus, at one-fold coverage, approximately 37% of a polynucleotide of size  $L$ , in nucleotides has not been sequenced. When 14 Mb of sequence has been generated, coverage is 5X for a 2.8 Mb and the unsequenced fraction drops to .0067 or 0.67%. 5X coverage of a 2.8 Mb sequence can be attained by sequencing approximately 17,000 random clones from both insert ends with an average sequence read length of 410 bp.

Similarly, the total gap length,  $G$ , is determined by the equation  $G = Le^{-m}$ , and the average gap size,  $g$ , follows the equation,  $g = L/n$ . Thus, 5X coverage leaves about 240 gaps averaging about 82 bp in size in a sequence of a polynucleotide 2.8 Mb long.

The treatment above is essentially that of Lander and Waterman, *Genomics* 2: 231 (1988).

## 2. Random Library Construction

In order to approximate the random model described above during actual sequencing, a nearly ideal library of cloned genomic fragments is required. The following library construction procedure was developed to achieve this end.

*Streptococcus pneumoniae* DNA is prepared by phenol extraction. A mixture containing 200  $\mu$ g DNA in 1.0 ml of 300 mM sodium acetate, 10 mM Tris-HCl, 1 mM Na-EDTA, 50% glycerol is processed through a nebulizer (IPI Medical Products) with a stream of nitrogen adjusted to 35 Kpa for 2 minutes. The sonicated DNA is ethanol precipitated and redissolved in 500  $\mu$ l TE buffer.

To create blunt-ends, a 100  $\mu$ l aliquot of the resuspended DNA is digested with 5 units of BAL31 nuclease (New England BioLabs) for 10 min at 30°C in 200  $\mu$ l BAL31 buffer. The digested DNA is phenol-extracted, ethanol-precipitated, redissolved in 100  $\mu$ l TE buffer, and then size-fractionated by electrophoresis through a 1.0% low melting temperature agarose gel. The section containing DNA fragments 1.6-2.0 kb in size is excised from the gel, and the LGT agarose is melted and the resulting solution is extracted with phenol to separate the agarose from the

DNA. DNA is ethanol precipitated and redissolved in 20  $\mu$ l of TE buffer for ligation to vector.

A two-step ligation procedure is used to produce a plasmid library with 97% inserts, of which >99% were single inserts. The first ligation mixture (50  $\mu$ l) contains 2  $\mu$ g of DNA fragments, 2  $\mu$ g pUC18 DNA (Pharmacia) cut with SmaI and dephosphorylated with bacterial alkaline phosphatase, and 10 units of T4 ligase (GIBCO/BRL) and is incubated at 14°C for 4 hr. The ligation mixture then is phenol extracted and ethanol precipitated, and the precipitated DNA is dissolved in 20  $\mu$ l TE buffer and electrophoresed on a 1.0% low melting agarose gel. Discrete bands in a ladder are visualized by ethidium bromide-staining and UV illumination and identified by size as insert (I), vector (v), v+I, v+2i, v+3i, etc. The portion of the gel containing v+I DNA is excised and the v+I DNA is recovered and resuspended into 20  $\mu$ l TE. The v+I DNA then is blunt-ended by T4 polymerase treatment for 5 min. at 37°C in a reaction mixture (50  $\mu$ l) containing the v+I linears, 500  $\mu$ M each of the 4 dNTPs, and 9 units of T4 polymerase (New England BioLabs), under recommended buffer conditions. After phenol extraction and ethanol precipitation the repaired v+I linears are dissolved in 20  $\mu$ l TE. The final ligation to produce circles is carried out in a 50  $\mu$ l reaction containing 5  $\mu$ l of v+I linears and 5 units of T4 ligase at 14°C overnight. After 10 min. at 70°C the following day, the reaction mixture is stored at -20°C.

This two-stage procedure results in a molecularly random collection of single-insert plasmid recombinants with minimal contamination from double-insert chimeras (<1%) or free vector (<3%).

Since deviation from randomness can arise from propagation the DNA in the host, *E. coli* host cells deficient in all recombination and restriction functions (A. Greener, *Strategies* 3 (1):5 (1990)) are used to prevent rearrangements, deletions, and loss of clones by restriction. Furthermore, transformed cells are plated directly on antibiotic diffusion plates to avoid the usual broth recovery phase which allows multiplication and selection of the most rapidly growing cells.

Plating is carried out as follows. A 100  $\mu$ l aliquot of Epicurian Coli SURE II Supercompetent Cells (Stratagene 200152) is thawed on ice and transferred to a chilled Falcon 2059 tube on ice. A 1.7  $\mu$ l aliquot of 1.42 M beta-mercaptoethanol is added to the aliquot of cells to a final concentration of 25 mM. Cells are incubated on ice for 10 min. A 1  $\mu$ l aliquot of the final ligation is added to the cells and incubated on ice for 30 min. The cells are heat pulsed for 30 sec. at 42°C and

placed back on ice for 2 min. The outgrowth period in liquid culture is eliminated from this protocol in order to minimize the preferential growth of any given transformed cell. Instead the transformation mixture is plated directly on a nutrient rich SOB plate containing a 5 ml bottom layer of SOB agar (5% SOB agar: 20 g tryptone, 5 g yeast extract, 0.5 g NaCl, 1.5% Difco Agar per liter of media). The 5 ml bottom layer is supplemented with 0.4 ml of 50 mg/ml ampicillin per 100 ml SOB agar. The 15 ml top layer of SOB agar is supplemented with 1 ml X-Gal (2%), 1 ml MgCl<sub>2</sub> (1 M), and 1 ml MgSO<sub>4</sub> /100 ml SOB agar. The 15 ml top layer is poured just prior to plating. Our titer is approximately 100 colonies/10  $\mu$ l aliquot of transformation.<sup>2</sup><sub>4</sub>

All colonies are picked for template preparation regardless of size. Thus, only clones lost due to "poison" DNA or deleterious gene products are deleted from the library, resulting in a slight increase in gap number over that expected.

### 3. Random DNA Sequencing

High quality double stranded DNA plasmid templates are prepared using a "boiling bead" method developed in collaboration with Advanced Genetic Technology Corp. (Gaithersburg, MD) (Adams *et al.*, *Science* 252:1651 (1991); Adams *et al.*, *Nature* 355:632 (1992)). Plasmid preparation is performed in a 96-well format for all stages of DNA preparation from bacterial growth through final DNA purification. Template concentration is determined using Hoechst Dye and a Millipore Cytofluor. DNA concentrations are not adjusted, but low-yielding templates are identified where possible and not sequenced.

Templates are also prepared from two *Streptococcus pneumoniae* lambda genomic libraries. An amplified library is constructed in the vector Lambda GEM-12 (Promega) and an unamplified library is constructed in Lambda DASH II (Stratagene). In particular, for the unamplified lambda library, *Streptococcus pneumoniae* DNA (> 100 kb) is partially digested in a reaction mixture (200  $\mu$ l) containing 50  $\mu$ g DNA, 1X Sau3AI buffer, 20 units Sau3AI for 6 min. at 23°C. The digested DNA was phenol-extracted and electrophoresed on a 0.5% low melting agarose gel at 2V/cm for 7 hours. Fragments from 15 to 25 kb are excised and recovered in a final volume of 6  $\mu$ l. One  $\mu$ l of fragments is used with 1  $\mu$ l of DASHII vector (Stratagene) in the recommended ligation reaction. One  $\mu$ l of the ligation mixture is used per packaging reaction following the recommended protocol with the Gigapack II XL Packaging Extract (Stratagene, #227711). Phage

are plated directly without amplification from the packaging mixture (after dilution with 500  $\mu$ l of recommended SM buffer and chloroform treatment). Yield is about  $2.5 \times 10^3$  pfu/ $\mu$ l. The amplified library is prepared essentially as above except the lambda GEM-12 vector is used. After packaging, about  $3.5 \times 10^4$  pfu are plated on the restrictive NM539 host. The lysate is harvested in 2 ml of SM buffer and stored frozen in 7% dimethylsulfoxide. The phage titer is approximately  $1 \times 10^9$  pfu/ml.

Liquid lysates (100  $\mu$ l) are prepared from randomly selected plaques (from the unamplified library) and template is prepared by long-range PCR using T7 and T3 vector-specific primers.

Sequencing reactions are carried out on plasmid and/or PCR templates using the AB Catalyst LabStation with Applied Biosystems PRISM Ready Reaction Dye Primer Cycle Sequencing Kits for the M13 forward (M13-21) and the M13 reverse (M13RP1) primers (Adams *et al.*, *Nature* 368:474 (1994)). Dye terminator sequencing reactions are carried out on the lambda templates on a Perkin-Elmer 9600 Thermocycler using the Applied Biosystems Ready Reaction Dye Terminator Cycle Sequencing kits. T7 and SP6 primers are used to sequence the ends of the inserts from the Lambda GEM-12 library and T7 and T3 primers are used to sequence the ends of the inserts from the Lambda DASH II library. Sequencing reactions are performed by eight individuals using an average of fourteen AB 373 DNA Sequencers per day. All sequencing reactions are analyzed using the Stretch modification of the AB 373, primarily using a 34 cm well-to-read distance. The overall sequencing success rate very approximately is about 85% for M13-21 and M13RP1 sequences and 65% for dye-terminator reactions. The average usable read length is 485 bp for M13-21 sequences, 445bp for M13RP1 sequences, and 375 bp for dye-terminator reactions.

Richards *et al.*, Chapter 28 in AUTOMATED DNA SEQUENCING AND ANALYSIS, M. D. Adams, C. Fields, J. C. Venter, Eds., Academic Press, London, (1994) described the value of using sequence from both ends of sequencing templates to facilitate ordering of contigs in shotgun assembly projects of lambda and cosmid clones. We balance the desirability of both-end sequencing (including the reduced cost of lower total number of templates) against shorter read-lengths for sequencing reactions performed with the M13RP1 (reverse) primer compared to the M13-21 (forward) primer. Approximately one-half of the templates are sequenced from both ends. Random reverse sequencing reactions are

done based on successful forward sequencing reactions. Some M13RP1 sequences are obtained in a semi-directed fashion: M13-21: sequences pointing outward at the ends of contigs are chosen for M13RP1 sequencing in an effort to specifically order contigs.

5

#### 4. Protocol for Automated Cycle Sequencing

The sequencing is carried out using ABI Catalyst robots and AB 373 Automated DNA Sequencers. The Catalyst robot is a publicly available sophisticated pipetting and temperature control robot which has been developed specifically for DNA sequencing reactions. The Catalyst combines pre-aliquoted templates and reaction mixes consisting of deoxy- and dideoxynucleotides, the thermostable Taq DNA polymerase, fluorescently-labelled sequencing primers, and reaction buffer. Reaction mixes and templates are combined in the wells of an aluminum 96-well thermocycling plate. Thirty consecutive cycles of linear amplification (i.e., one primer synthesis) steps are performed including denaturation, annealing of primer and template, and extension; i.e., DNA synthesis. A heated lid with rubber gaskets on the thermocycling plate prevents evaporation without the need for an oil overlay.

Two sequencing protocols are used: one for dye-labelled primers and a second for dye-labelled dideoxy chain terminators. The shotgun sequencing involves use of four dye-labelled sequencing primers, one for each of the four terminator nucleotide. Each dye-primer is labelled with a different fluorescent dye, permitting the four individual reactions to be combined into one lane of the 373 DNA Sequencer for electrophoresis, detection, and base-calling. ABI currently supplies pre-mixed reaction mixes in bulk packages containing all the necessary non-template reagents for sequencing. Sequencing can be done with both plasmid and PCR-generated templates with both dye-primers and dye-terminators with approximately equal fidelity, although plasmid templates generally give longer usable sequences.

Thirty-two reactions are loaded per AB373 Sequencer each day, for a total of 960 samples. Electrophoresis is run overnight following the manufacturer's protocols, and the data is collected for twelve hours. Following electrophoresis and fluorescence detection, the ABI 373 performs automatic lane tracking and base-calling. The lane-tracking is confirmed visually. Each sequence electropherogram (or fluorescence lane trace) is inspected visually and assessed for quality. Trailing

sequences of low quality are removed and the sequence itself is loaded via software to a Sybase database (archived daily to 8mm tape). Leading vector polylinker sequence is removed automatically by a software program. Average edited lengths of sequences from the standard ABI 373 are around 400 bp and depend mostly on the quality of the template used for the sequencing reaction. ABI 373 Sequencers converted to Stretch Liners provide a longer electrophoresis path prior to fluorescence detection and increase the average number of usable bases to 500-600 bp.

## 10           **INFORMATICS**

### 1. Data Management

A number of information management systems for a large-scale sequencing lab have been developed. (For review see, for instance, Kerlavage *et al.*, *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press, Washington D. C., 585 (1993)) The system used to collect and assemble the sequence data was developed using the Sybase relational database management system and was designed to automate data flow wherever possible and to reduce user error. The database stores and correlates all information collected during the entire operation from template preparation to final analysis of the genome. Because the raw output of the ABI 373 Sequencers was based on a Macintosh platform and the data management system chosen was based on a Unix platform, it was necessary to design and implement a variety of multi-user, client-server applications which allow the raw data as well as analysis results to flow seamlessly into the database with a minimum of user effort.

25

### 2. Assembly

An assembly engine (TIGR Assembler) developed for the rapid and accurate assembly of thousands of sequence fragments was employed to generate contigs. The TIGR assembler simultaneously clusters and assembles fragments of the genome. In order to obtain the speed necessary to assemble more than  $10^4$  fragments, the algorithm builds a hash table of 12 bp oligonucleotide subsequences to generate a list of potential sequence fragment overlaps. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Beginning with a single seed sequence fragment, TIGR Assembler extends the current contig by attempting to add the best matching

35

fragment based on oligonucleotide content. The contig and candidate fragment are aligned using a modified version of the Smith-Waterman algorithm which provides for optimal gapped alignments (Waterman, M. S., *Methods in Enzymology* 164:765 (1988)). The contig is extended by the fragment only if strict criteria for the quality of the match are met. The match criteria include the minimum length of overlap, the maximum length of an unmatched end, and the minimum percentage match. These criteria are automatically lowered by the algorithm in regions of minimal coverage and raised in regions with a possible repetitive element. The number of potential overlaps for each fragment determines which fragments are likely to fall into repetitive elements. Fragments representing the boundaries of repetitive elements and potentially chimeric fragments are often rejected based on partial mismatches at the ends of alignments and excluded from the current contig. TIGR Assembler is designed to take advantage of clone size information coupled with sequencing from both ends of each template. It enforces the constraint that sequence fragments from two ends of the same template point toward one another in the contig and are located within a certain range of base pairs (definable for each clone based on the known clone size range for a given library).

The process resulted in 391 contigs as represented by SEQ ID NOs:1-391.

### 3. Identifying Genes

The predicted coding regions of the *Streptococcus pneumoniae* genome were initially defined with the program GeneMark, which finds ORFs using a probabilistic classification technique. The predicted coding region sequences were used in searches against a database of all nucleotide sequences from GenBank (October, 1997), using the BLASTN search method to identify overlaps of 50 or more nucleotides with at least a 95% identity. Those ORFs with nucleotide sequence matches are shown in Table 1. The ORFs without such matches were translated to protein sequences and compared to a non-redundant database of known proteins generated by combining the Swiss-prot, PIR and GenPept databases. ORFs that matched a database protein with BLASTP probability less than or equal to 0.01 are shown in Table 2. The table also lists assigned functions based on the closest match in the databases. ORFs that did not match protein or nucleotide sequences in the databases at these levels are shown in Table 3.



## ILLUSTRATIVE APPLICATIONS

### 1. Production of an Antibody to a *Streptococcus pneumoniae* Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells using any one of the methods known in the art. The protein can also be produced in a recombinant prokaryotic expression system, such as *E. coli*, or can be chemically synthesized. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows.

### 2. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., *Nature* 256:495 (1975) or modifications of the methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and modified methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. *et al.*, *Basic Methods in Molecular Biology*, Elsevier, New York. Section 21-2 (1989).

### 3. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. *et al.*, *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. *et al.*, Chap. 19 in: *Handbook of Experimental Immunology*, Wier, D., ed, Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12M). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, second edition, Rose and Friedman, eds., Amer. Soc. For Microbiology, Washington, D. C. (1980)

Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi- quantitatively or qualitatively to identify the presence of antigen in a biological sample. In addition, antibodies are useful in various animal models of pneumococcal disease as a means of evaluating the protein used to make the antibody as a potential vaccine target or as a means of evaluating the antibody as a potential immunotherapeutic or immunoprophylactic reagent.

#### 4. Preparation of PCR Primers and Amplification of DNA

Various fragments of the *Streptococcus pneumoniae* genome, such as those of Tables 1-3 and SEQ ID NOS:1-391 can be used, in accordance with the present invention, to prepare PCR primers for a variety of uses. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. When selecting a primer sequence, it is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

#### 5. Gene expression from DNA Sequences Corresponding to ORFs

A fragment of the *Streptococcus pneumoniae* genome provided in Tables 1-3 is introduced into an expression vector using conventional technology. Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield *et al.*, U. S. Patent No. 5,082,767, incorporated herein by this reference.

The following is provided as one exemplary method to generate polypeptide(s) from cloned ORFs of the *Streptococcus pneumoniae* genome fragment. Bacterial ORFs generally lack a poly A addition signal. The addition signal sequence can be added to the construct by, for example, splicing out the poly A addition sequence from pSG5 (Stratagene)—using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene) for use in eukaryotic expression systems. pXT1 contains the LTRs and a portion of the gag gene of Moloney Murine Leukemia Virus. The positions of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The *Streptococcus pneumoniae* DNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the *Streptococcus pneumoniae* DNA and containing restriction endonuclease sequences for PstI incorporated into the 5' primer and BglII at the 5' end of the corresponding *Streptococcus pneumoniae* DNA 3' primer, taking care to ensure that the *Streptococcus pneumoniae* DNA is positioned such that its followed with the poly A addition sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with BglII, purified and ligated to pXT1, now containing a poly A addition sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600 ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface. Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted *Streptococcus pneumoniae* DNA sequence are injected into mice to generate antibody to the polypeptide encoded by the *Streptococcus pneumoniae* DNA.

Alternatively and if antibody production ~~is not~~ possible, the *Streptococcus pneumoniae* DNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as, for example, a globin fusion. Antibody to the globin moiety then is used to purify the chimeric protein. Corresponding protease cleavage sites are engineered between the globin moiety and the polypeptide encoded by the *Streptococcus pneumoniae* DNA so that the latter may be freed from the formed by simple protease digestion. One useful expression vector for generating globin chimerics is pSG5 (Stratagene). This vector encodes a rabbit globin. Intron II of the rabbit globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, cited elsewhere herein, and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptides of the invention also may be produced using *in vitro* translation systems such as *in vitro* Express<sup>TM</sup> Translation Kit (Stratagene).

While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

All patents, patent applications and publications referred to above are hereby incorporated by reference.

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
1	1	437	1003	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	92	200	567
2	5	6169	5720	gb U04047	Streptococcus pneumoniae SSZ dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
2	6	6592	6167	emb Z83335 SP28	S. pneumoniae dexB, cap1(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	98	426	426
3	11	9770	9147	emb Z83335 SP28	S. pneumoniae dexB, cap1(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	624	624
3	12	10489	9671	emb Z83335 SP28	S. pneumoniae dexB, cap1(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	819	819
3	13	11546	12019	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	474	474
3	14	12017	13375	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1359	1359
3	15	13421	14338	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	918	918
3	16	14329	15171	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	843	843
3	17	15112	17282	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	2151	2151
3	18	17267	18397	gb U43526	Streptococcus pneumoniae neuraminidase B (nanB) gene, complete cds, and neuraminidase (nanA) gene, partial cds	99	1069	1131
4	1	46	1188	emb Y11463 SPUN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1143	1163
4	2	1198	2529	emb Y11463 SPUN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	876	1332
5	7	11297	11473	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	82	175	177
6	7	7125	7364	emb Z77726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	93	238	240
6	8	7322	7570	emb Z77726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (966 bp)	95	160	249
6	9	7533	7985	emb Z77726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (966 bp)	99	453	453
6	23	20197	19711	emb Z83335 SP28	S. pneumoniae dexB, cap1(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	96	465	465
7	10	8105	7682	emb Z83335 SP28	S. pneumoniae dexB, cap1(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	624	624

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
7	11	9024	8206	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	95	819	819
10	13	9304	8078	gb L29333	Streptococcus pneumoniae methyl transferase (mtr) gene cluster, complete cds	93	513	1227
11	2	548	919	emb Z79691 SOOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	99	316	372
11	3	892	1980	emb Z79691 SUOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	99	1089	1089
11	5	3040	3477	emb Z79691 SOOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	99	259	438
11	6	3480	3247	emb Z79691 SOOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	99	234	234
11	7	3601	4557	emb Z79691 SOOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	98	957	957
11	8	4506	4886	emb Z79691 SOOR	S.pneumoniae yor(A,B,C,D,E), <i>ftsL</i> , <i>pbpX</i> and <i>regR</i> genes	99	381	381
11	9	4884	7142	emb X16367 SPPB	Streptococcus pneumoniae <i>pbpX</i> gene for penicillin binding protein 2X	99	2259	2259
11	10	7132	8124	emb X16367 SPPB	Streptococcus pneumoniae <i>pbpX</i> gene for penicillin binding protein 2X	98	70	993
13	1	53	1126	gb H31296	S.pneumoniae <i>recP</i> gene, complete cds	99	437	1074
14	3	1837	2148	emb Z83335 SP28	S.pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	87	96	312
14	4	2518	2108	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase ( <i>putC</i> ) genes, complete cds	98	411	411
15	9	8942	8511	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cps19FABCDGHIJKLMNO) genes, complete cds, and <i>alia</i> gene, partial cds	89	340	432
17	7	3910	3458	emb Z77726 SP15	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	453	453
17	8	4304	3873	emb Z77727 SP15	S.pneumoniae DNA for insertion sequence IS1318 (823 bp)	96	382	432
19	1	41	529	emb X94909 SP1G	S.pneumoniae <i>iga</i> gene	75	368	489
19	2	554	757	gb L07752	Streptococcus pneumoniae attachment site ( <i>attB</i> ), DNA sequence	99	167	204
19	3	946	1827	gb L07752	Streptococcus pneumoniae attachment site ( <i>attB</i> ), DNA sequence	94	100	882
20	1	937	182	gb U3315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor ( <i>comC</i> ), histidine protein kinase ( <i>comB</i> ) and response regulator ( <i>comE</i> ) genes, complete cds, <i>trnA</i> -Arg and <i>trnA</i> -Gln genes	99	756	756
20	2	2271	931	gb U3315	Streptococcus pneumoniae orfL gene, partial cds, competence stimulating peptide precursor ( <i>comC</i> ), histidine protein kinase ( <i>comB</i> ) and response regulator ( <i>comE</i> ) genes, complete cds, <i>trnA</i> -Arg and <i>trnA</i> -Gln genes	98	3341	3341

TABLE 1  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
20	1	3175	2684	gb U76218	Streptococcus pneumoniae Competence stimulating peptide precursor ComC (comC), histidine kinase homolog ComD (comD), and response regulator homolog ComE (comE) genes, complete cds	99	492	492
20	4	3322	4527	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	99	1206	1206
20	5	4573	5343	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	99	771	771
20	6	5532	6917	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	99	1386	1386
20	7	6995	8212	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	99	1218	1218
20	8	8214	8471	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	98	258	258
20	9	8534	9670	gb AF000658	Streptococcus pneumoniae R801 tRNA-Arg gene, partial sequence, and putative serine protease (spitral), SPSPoj (spspoj), initiator protein (spdnal) and beta subunit of DNA polymerase III (spdnal) genes, complete cds	99	1134	1137
22	14	11887	12267	emb Z77726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	99	226	381
22	15	12708	12356	emb Z77727 SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	453
22	16	13165	12662	emb Z77726 SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	98	504	504
22	23	18398	18910	emb Z86112 SP28	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	95	463	513
22	24	18829	19299	emb Z86112 SP28	S. pneumoniae genes encoding galacturonosyl transferase and transposase and insertion sequence IS1515	99	443	471
23	5	5624	4203	emb X52474 SPPL	S. pneumoniae ply gene for pneumolysin	99	1422	1422
23	6	6063	5629	gb H17717	S. pneumoniae pneumolysin gene, complete cds	98	197	435
26	1	5500	2	emb X94909 SP1C	S. pneumoniae iga gene	87	3487	5499
26	2	5823	5584	gb U7687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	99	151	240
26	3	6878	5685	gb U7687	Streptococcus pneumoniae immunoglobulin A1 protease (iga) gene, complete cds	100	50	1194



TABLE I

S pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	match gene name	percent ident	HSP nt length	ORF nt length
26	8	14498	14854	emb 283335 SP28	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	99	338	357
26	9	14763	14924	emb 283335 SP28	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	100	94	162
26	10	14922	15171	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
28	1	80	505	emb 283335 SP28	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	99	426	426
28	2	503	952	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	450	450
28	3	780	1298	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	181	519
34	1	207	1523	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	99	1317	1317
34	2	1477	2367	gb U08611	Streptococcus pneumoniae maltose/maltodextrin uptake (malX) and two maltodextrin permease (malC and malD) genes, complete cds	96	795	891
34	3	2593	3420	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	446	828
34	4	2790	2647	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	98	137	144
34	5	3418	4416	gb L21856	Streptococcus pneumoniae malA gene, complete cds; malR gene, complete cds	96	999	999
34	9	7764	7507	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	93	201	258
34	16	10562	10257	emb X63602 SP80	S pneumoniae mmsA-Box	92	238	306
35	4	1176	1439	emb 283335 SP28	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	87	248	264
35	5	1458	1961	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsI9fABCDEFHIJKLMNO) genes, complete cds, and <i>alia</i> gene, partial cds	98	264	504
35	17	16172	15477	emb X85787 SPCP	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	97	696	696
35	18	16961	16170	emb 283335 SP28	S pneumoniae dexB, capsIIA, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and <i>alia</i> gene	86	792	792
35	19	17620	16871	gb U09239	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsI9fABCDEFHIJKLMNO) genes, complete cds, and <i>alia</i> gene, partial cds	83	750	750

**S. pneumoniae - Coding regions containing known sequences**

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
35	20	19061	17604	emb X85787 SPCV	S.pneumoniae dexB, cps14A, cps14B, cps14C, cps14D, cps14E, cps14F, cps14G, cps14H, cps14I, cps14J, cps14K, cps14L, csa genes	94	1458	1458
36	19	18960	18352	gb u040786	Streptococcus pneumoniae surface antigen A variant precursor (psaA) and 18 kDa protein genes, complete cds, and ORF1 gene, partial cds	99	609	609
36	20	19934	18966	gb u053509	Streptococcus pneumoniae surface adhesin A precursor (psaA) gene, complete cds	99	969	969
37	1	2743	179	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	2565	2565
37	2	2985	2824	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	162	162
37	3	5034	3070	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	1965	1965
37	4	5134	5790	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	99	657	657
37	5	6171	5833	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	96	339	339
38	19	12969	13268	gb M28679	S.pneumoniae promoter region DNA	100	64	300
39	2	1256	2137	gb u041735	Streptococcus pneumoniae peptidic methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	882	882
39	3	2405	3370	gb u041735	Streptococcus pneumoniae peptidic methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	99	966	966
40	9	5253	7208	gb M29686	S.pneumoniae mismatch repair (hmx8) gene, complete cds	99	1956	1956
41	1	3	1037	emb Z17307 SPRE	S.pneumoniae recA gene encoding RecA	99	1027	1035
41	2	1328	2713	emb Z34303 SPCI	Streptococcus pneumoniae cln operon encoding the clnA, recA, difN, lytA genes, and downstream sequences	99	1386	1386
41	3	3083	4045	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	99	963	963
41	4	3272	3096	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	177	177
41	5	3603	3860	gb M13812	S.pneumoniae autolysin (lytA) gene, complete cds	100	258	258
41	6	4755	5162	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	408	408
41	7	5270	5716	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	447	447
41	8	6112	6918	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	431	807
41	9	6916	7119	gb L36660	Streptococcus pneumoniae ORF, complete cds	100	204	204
41	10	7082	7660	gb L36660	Streptococcus pneumoniae ORF, complete cds	97	552	579
41	11	7680	7979	gb L36660	Streptococcus pneumoniae ORF, complete cds	98	81	300
41	12	9169	8717	emb Z77727 SPIS	S.pneumoniae DNA for insertion sequence IS1310 (823 bp)	97	351	453

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
41	13	9533	9132	emb Z77725 SP1S	S. pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
41	14	9659	9475	emb Z82001 SP28	S. pneumoniae pcps gene and open reading frames	100	189	195
44	5	7150	7555	emb Z82001 SP28	S. pneumoniae pcps gene and open reading frames	99	366	366
44	6	8059	7607	emb Z77726 SP1S	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	97	453	453
44	7	8423	8022	emb Z77725 SP1S	S. pneumoniae DNA for insertion sequence IS1381 (966 bp)	95	160	402
44	8	8559	8365	emb Z82001 SP28	S. pneumoniae pcps gene and open reading frames	100	189	195
48	9	6480	4687	gb U139074	Streptococcus pneumoniae pyruvate oxidase (spox) gene, complete cds	99	1794	1794
49	2	231	2603	gb U20561	Streptococcus pneumoniae Exp7 gene, partial cds	100	216	2373
51	6	2407	2156	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	242	252
53	7	2566	2405	emb Z83335 SP28	S. pneumoniae dexB, capl(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and allia gene	100	94	162
53	8	2831	2475	emb Z83335 SP28	S. pneumoniae dexB, capl(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and allia gene	99	338	357
54	11	12409	11105	emb Z83335 SP28	S. pneumoniae dexB, capl(A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and allia gene	67	591	1305
55	22	20488	19949	emb Z83379 HS28	S. pneumoniae dfr gene (isolate 92)	99	540	540
61	11	11864	9900	emb Z16082 PNAL	Streptococcus pneumoniae allia gene	98	1965	1965
63	1	3	239	gb H18729	S. pneumoniae mismatch repair protein (hexA) gene, complete cds	100	237	237
63	2	233	2611	gb H18729	S. pneumoniae mismatch repair protein (hexA) gene, complete cds	99	2330	2379
63	3	2557	2823	gb H18729	S. pneumoniae mismatch repair protein (hexA) gene, complete cds	99	266	267
63	4	2958	4664	gb H18729	S. pneumoniae mismatch repair protein (hexA) gene, complete cds	95	69	1707
67	6	3770	3399	gb U20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	96	372	372
67	7	7161	4171	gb U20670	Streptococcus pneumoniae hyaluronidase gene, complete cds	99	2938	2991
70	1	1	702	gb H14340	S. pneumoniae DpnI gene region encoding dpcC and dpcD, complete cds	100	693	702
70	2	678	1160	gb H14340	S. pneumoniae DpnI gene region encoding dpcC and dpcD, complete cds	100	483	483
70	3	2490	1210	gb H14339	S. pneumoniae DpnII gene region encoding dpcM, dpcN, dpcO, complete cds	98	462	1281
70	7	4230	4424	gb U04234	S. pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	147	195
70	8	5197	4316	gb U04234	S. pneumoniae exodeoxyribonuclease (exoA) gene, complete cds	99	881	882

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
70	13	8108	9874	gb L20562	Streptococcus pneumoniae Exp8 gene, partial cds	93	234	1767
71	22	27964	28341	emb X63602 SP80	S pneumoniae masA-Box	93	233	378
72	5	4607	3552	emb 226850 SPAT	S pneumoniae (M223) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	1056
73	1	471	133	emb X63602 SP80	S pneumoniae masA-Box	91	193	339
73	3	3658	977	gb J04479	S pneumoniae DNA polymerase I (polA) gene, complete cds	99	2682	2682
73	8	4864	5379	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	318	516
77	3	2622	1999	emb 283335 SP28	S pneumoniae dexB, cap11A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	624	624
77	4	3341	2523	emb 283335 SP28	S pneumoniae dexB, cap11A, B, C, D, E, F, G, H, I, J, K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	819	819
78	1	341	3	emb X77249 SPR6	S pneumoniae (R6) clar/clah genes	99	339	339
78	2	1095	325	emb X77249 SPR6	S pneumoniae (R6) clar/clah genes	99	771	771
82	10	11436	10816	gb U90721	Streptococcus pneumoniae signal peptidase I (spi) gene, complete cds	97	621	621
82	11	12402	11434	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	98	953	969
82	12	12381	12704	gb U93576	Streptococcus pneumoniae ribonuclease HII (rnhB) gene, complete cds	100	51	324
83	8	3212	3550	emb 277727 SP18	S pneumoniae RNA for insertion sequence IS1318 (823 bp)	97	290	339
83	10	4662	6851	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	2190	2190
83	11	6849	8213	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	1365	1365
83	12	8216	9090	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	855	855
83	13	9283	13017	gb L15190	Streptococcus pneumoniae SAICAR synthetase (purC) gene, complete cds	100	107	3735
83	23	22147	23313	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (stxH) gene, complete cds	98	218	1167
83	24	23268	23450	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (stxH) gene, complete cds	98	172	183
83	25	27523	23505	gb L36923	Streptococcus pneumoniae beta-N-acetylhexosaminidase (stxH) gene, complete cds	99	3826	4023

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	RSP nt length	ORF nt length
83	26	28472	27771	[gb UJ6923]	Streptococcus pneumoniae beta-N-acetylhexosaminidase (stxH) gene, complete cds	99	416	702
84	3	4554	6173	[emb 283335]SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-thiamnose biosynthesis genes and aliA gene	98	697	1620
87	6	5951	5316	[emb 27725]SP15	S. pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	439	636
88	5	2957	3511	[gb HJ6180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	555	555
88	6	3466	4269	[gb HJ6180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	804	804
89	13	9878	10093	[gb HJ6180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	97	211	216
89	14	10062	10412	[emb 283335]SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-thiamnose biosynthesis genes and aliA gene	97	335	351
91	10	5103	4941	[emb X63602]SP80	S. pneumoniae msaA-Box	89	237	363
97	4	1708	1520	[gb UJ41735]	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	91	140	189
99	1	89	700	[emb 283335]SP28	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-thiamnose biosynthesis genes and aliA gene	93	592	612
99	2	1773	775	[emb X17337]SPAH	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	99	998	999
99	3	2794	1712	[emb X17337]SPAH	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	99	1083	1083
99	4	3732	2788	[emb X17337]SPAH	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	100	945	945
99	5	5249	3714	[emb X17337]SPAH	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	100	1536	1536
99	6	1262	5277	[emb X17337]SPAH	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	99	1986	1986
101	1	216	1538	[emb X54225]SP2H	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	146	1323
101	2	1492	1719	[emb X54225]SP2H	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	99	228	228
101	3	1694	1855	[emb X54225]SP2H	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	162	162
101	4	1701	2582	[emb X54225]SP2H	S. pneumoniae epuA and endA genes for 7 kDa protein and membrane endonuclease	100	882	882
103	7	5556	5041	[emb 295914]SP29	Streptococcus pneumoniae sodA gene	100	396	516
104	2	1347	1556	[emb 27727]SP15	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	83	206	210

TABLE 1

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
105	5	5381	5028	emb Z67739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	98	353	354
105	6	6089	5379	emb Z67739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	98	84	711
107	4	2785	1880	emb X16022 SPPE	S. pneumoniae penA gene	98	72	906
107	5	2913	4988	emb X16022 SPPE	S. pneumoniae penA gene	99	1692	2076
107	6	4981	5595	emb X13116 SPPE	Streptococcus pneumoniae penA gene for penicillin binding protein 2b lacking N-term. (penicillin resistant strain)	91	107	615
108	9	9068	8718	emb Z67739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	95	342	351
108	12	11108	10922	emb Z67739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	99	199	387
109	3	2768	2241	emb Z77725 SPIS	S. pneumoniae DNA for insertion sequence IS1381 (966 bp)	96	61	528
109	4	2688	2855	emb Z77726 SPIS	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	148	168
109	5	2862	3269	emb Z77727 SPIS	S. pneumoniae DNA for insertion sequence IS1318 (823 bp)	97	353	408
109	6	5320	3584	gb H18729	S. pneumoniae mismatch repair protein (lexA) gene, complete cds	100	371	1737
113	1	431	3	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	429	429
113	10	9788	8532	emb X99400 SPDA	S. pneumoniae dacA gene and ORF	99	1257	1257
113	11	9870	10985	emb X99400 SPDA	S. pneumoniae dacA gene and ORF	99	1116	1116
114	3	2510	2030	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	95	481	501
115	11	11130	10932	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	97	372	372
117	1	897	3302	emb X72967 SPNA	S. pneumoniae nanA gene	99	2402	2406
117	2	3277	3831	emb X72967 SPNA	S. pneumoniae nanA gene	99	237	555
117	3	4327	3899	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	429	429
121	2	1369	1941	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	202	573
121	3	2412	4251	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and DnaJ (dnaJ) gene, partial cds	99	1842	1842
122	8	5066	5587	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	64	451	522

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
125	1	1811	189	[gb U16180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (puIC) genes, complete cds	92	99	1623
128	15	12496	11204	[emb Z81335]SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiaminose biosynthesis genes and allA gene	91	705	1293
134	1	1	492	[emb Y10818]SPV1	S. pneumoniae spsA gene	99	203	492
134	2	556	2652	[gb AF019904]	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	86	685	2097
134	3	1160	837	[emb Y10818]SPV1	S. pneumoniae spsA gene	86	324	324
134	4	3952	2882	[gb AF019904]	Streptococcus pneumoniae choline binding protein A (cbpA) gene, partial cds	98	215	1071
134	8	7992	9848	[gb U12567]	Streptococcus pneumoniae p13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	285	1857
134	9	9846	10622	[gb U12567]	Streptococcus pneumoniae p13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	99	570	777
134	10	10805	11122	[gb U12567]	Streptococcus pneumoniae p13 glycerol-3-phosphate dehydrogenase (glpD) gene, partial cds, and glycerol uptake facilitator (glpF) and ORF3 genes, complete cds	100	318	318
137	13	7970	8443	[gb U09219]	Streptococcus pneumoniae type 19F capsular polysaccharide biosynthesis operon, (cpsY)(ABCEFGHIJXLMNO) genes, complete cds, and allA gene, partial cds	90	420	474
137	14	8590	8775	[emb Z81335]SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiaminose biosynthesis genes and allA gene	94	174	186
137	15	8773	8967	[emb Z81335]SP28	S. pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K, genes, dTDP-thiaminose biosynthesis genes and allA gene	98	195	195
137	16	9223	9687	[emb Z17726]SP15	S. pneumoniae DNA for insertion sequence IS1318 (1372 bp)	96	446	465
137	17	9641	10051	[emb Z17727]SP15	S. pneumoniae DNA for insertion sequence IS1318 (822 bp)	96	293	411
139	10	12998	12702	[emb X63602]SPB0	S. pneumoniae msaA-Box	90	234	297
141	8	7805	8938	[emb Z49988]SPM	Streptococcus pneumoniae msaA gene	99	338	1134
141	9	8916	10972	[emb Z49988]SPM	Streptococcus pneumoniae msaA gene	99	2037	2037
141	10	11472	12467	[emb Z49988]SPM	Streptococcus pneumoniae msaA gene	100	76	996
142	2	257	814	[gb M80215]	Streptococcus pneumoniae uva402 protein gene, complete cds	98	174	558
142	3	787	957	[gb M80215]	Streptococcus pneumoniae uva402 protein gene, complete cds	100	142	171
142	4	980	3022	[gb M80215]	Streptococcus pneumoniae uva402 protein gene, complete cds	95	1997	2043

TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
142	5	3020	3595	gb H80215	Streptococcus pneumoniae uvs402 protein gene, complete cds	100	153	576
145	1	1	219	emb Z35135 SPAL	S. pneumoniae alia gene for amIA-like gene A	97	185	219
145	2	171	1994	gb U20556	Streptococcus pneumoniae p1pA gene, partial cds	99	1811	1824
145	3	2287	7599	emb Z47210 SPDE	S. pneumoniae dexB, cap3A, cap3B and cap3C genes and orfA	99	1052	5313
145	4	9934	7766	gb H90527	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	2169	2169
145	5	10488	9922	gb H90527	Streptococcus pneumoniae penicillin-binding protein (ponA) gene, complete cds	99	512	567
146	1	159	4	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	98	156	156
146	2	344	90	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	98	255	255
146	16	11795	10794	emb Z82002 SP28	S. pneumoniae pcgB and pcgC genes	85	276	1002
147	11	10678	10202	emb Z21702 SPUH	S. pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	98	477	477
147	12	11318	10676	emb Z21702 SPUH	S. pneumoniae ung gene and mutX genes encoding uracil-DNA glycosylase and 8-oxodGTP nucleoside triphosphatase	99	663	663
148	12	9009	8815	gb U41735	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	90	180	195
156	4	1154	1402	emb X61602 SP80	S. pneumoniae mmsA-Box	94	185	249
159	11	9048	8521	gb H16180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	98	526	528
160	1	1	147	emb Z26851 SPAT	S. pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	100	142	147
160	2	179	898	emb Z26851 SPAT	S. pneumoniae (R6) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	99	720	720
160	3	906	1406	emb Z26850 SPAT	S. pneumoniae (H222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	95	501	501
160	4	1373	1942	emb Z26850 SPAT	S. pneumoniae (H222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	87	306	570
161	1	1	984	emb X77249 SPR6	S. pneumoniae (R6) clar/clah genes	99	984	984
161	7	6910	7497	emb X81917 SPCY	S. pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	99	437	588
161	8	7443	9386	emb X81917 SPCY	S. pneumoniae orf1gyrB and gyrB gene encoding DNA gyrase B subunit	98	1912	1944
163	1	2	2155	gb U20559	Streptococcus pneumoniae exp5 gene, partial cds	98	327	2154



TABLE I

S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	Percent Ident	HSP nt length	ORF nt length
165	1	32	1618	gb J01796	S.pneumoniae malX and malM genes encoding membrane protein and amyloamylase, complete cds, and malP gene encoding phosphorylase	99	1587	1587
165	2	1608	1902	gb J01796	S.pneumoniae malX and malM genes encoding membrane protein and amyloamylase, complete cds, and malP gene encoding phosphorylase	100	280	2295
166	1	378	4	emb Y11463 SPUN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF's	100	375	375
166	2	1507	320	emb Y11463 SPUN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	1188	1188
166	3	3240	1432	emb Y11463 SPUN	Streptococcus pneumoniae dnaG, rpoD, cpoA genes and ORF3 and ORF5	99	563	1809
167	1	1077	328	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	94	155	750
167	2	1844	999	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	98	405	846
167	3	2714	1842	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	97	604	873
167	4	3399	2641	emb Z71552 SPAD	Streptococcus pneumoniae adcCBA operon	99	703	759
168	1	1	2359	gb L20558	Streptococcus pneumoniae Exp4 gene, partial cds	99	282	2259
170	110	7338	7685	emb Z77726 SP15	S.pneumoniae DNA for insertion sequence IS1318 (1372 bp)	95	315	348
172	6	2462	4981	gb U47625	Streptococcus pneumoniae formate acetyltransferase (exp72) gene, partial cds	97	365	2520
175	1	373	20	gb H36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	89	351	354
175	4	1843	1621	emb Z47210 SPDE	S.pneumoniae dexB, cap1A, cap1B and cap1C genes and orf's	95	89	1779
176	5	3984	2980	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	100	573	1005
178	1	3	425	emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	423	423
179	1	426	70	emb Z83335 SP28	S.pneumoniae dexB, cap1A, B, C, D, E, F, G, H, I, J, K genes, JUDP-rhamnose biosynthesis genes and allA gene	99	338	357
180	3	3084	1855	emb X95718 SPGY	S.pneumoniae gyrA gene	99	381	1210
186	1	714	4	emb Z79691 SOOR	S.pneumoniae ynf(A, B, C, D, E), ftsL, pbpX and regR genes	98	59	711
186	2	2254	608	emb Z79691 SOOR	S.pneumoniae ynf(A, B, C, D, E), ftsL, pbpX and regR genes	98	315	1647
186	3	707	880	emb Z79691 SOOR	S.pneumoniae ynf(A, B, C, D, E), ftsL, pbpX and regR genes	98	174	174
189	1	2	259	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	258	258
189	2	600	185	gb U72720	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	98	204	216

TABLE I

S. pneumoniae - coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
189	3	1018	851	[gb U272720]	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	168	168
189	4	1012	2154	[gb U272720]	Streptococcus pneumoniae heat shock protein 70 (dnaK) gene, complete cds and dnaJ (dnaJ) gene, partial cds	99	1062	1143
191	9	7829	7524	[emb X63602 SP80]	S. pneumoniae msaA-Box	95	234	306
194	1	1	729	[gb M16180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	91	728	729
199	2	1117	881	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	96	211	237
199	4	1499	1762	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	89	248	264
199	5	1781	2284	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	98	504	504
203	1	1977	337	[gb U20563]	Streptococcus pneumoniae Exp-1 gene, partial cds	99	342	1641
204	1	1145	3	[gb L16131]	Streptococcus pneumoniae exp-10 gene, complete cds, recA gene, 5' end	99	1143	1143
208	1	59	2296	[gb U89711]	Streptococcus pneumoniae pneumococcal surface protein A PspA (pspA) gene, complete cds	90	471	2238
213	3	2455	2123	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	96	332	333
216	1	368	12	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	99	338	357
216	3	2650	2327	[gb M28678]	S. pneumoniae promoter sequence DNA	98	86	324
222	1	417	4	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	94	414	414
227	3	5266	4238	[emb AJ000336 SP]	Streptococcus pneumoniae ldh gene	99	1029	1029
239	1	1	804	[gb M1296]	S. pneumoniae recP gene, complete cds	95	484	804
247	3	1625	1807	[gb M16180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	94	178	183
249	3	921	1364	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	94	443	444
253	1	362	3	[gb M16180]	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (purC) genes, complete cds	99	360	360
253	5	1238	2050	[emb Z83335 SP28]	S. pneumoniae dexB, cap1(A,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and allia gene	95	420	813

TABLE I  
S. pneumoniae - Coding regions containing known sequences

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
253	6	2069	2572	emb 283335 SP28	S. pneumoniae dxB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and allA gene	97	504	504
255	1	3	800	emb 282002 SP28	S. pneumoniae pcgB and pcgC genes	97	511	798
255	2	798	1841	emb 282002 SP28	S. pneumoniae pcgB and pcgC genes	97	672	1044
255	3	2493	1969	emb 267739 SPPA	S. pneumoniae parC, parE and transposase genes and unknown orf	92	435	525
257	2	985	770	emb X17337 SPAM	Streptococcus pneumoniae am1 locus conferring aminopterin resistance	96	117	216
257	3	1245	907	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (putC) genes, complete cds	97	339	339
267	2	495	1208	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	95	84	714
267	3	1291	2277	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	97	755	987
267	4	2261	3601	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	98	1341	1341
267	5	3561	4136	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	576	576
267	6	4164	4949	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	99	748	786
267	7	5544	5140	gb U16156	Streptococcus pneumoniae dihydropteroate synthase (sulA), dihydrofolate synthetase (sulB), guanosine triphosphate cyclohydrolase (sulC), aldolase-pyrophosphokinase (sulD) genes, complete cds	100	186	405
268	4	1793	1990	emb X63602 SP80	S. pneumoniae masA-Box	89	194	198
271	1	562	104	gb M29686	S. pneumoniae mismatch repair (hexB) gene, complete cds	93	160	459
291	1	75	524	gb U04047	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, complete cds	96	450	450
291	2	1001	525	emb 283335 SP28	S. pneumoniae dxB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and allA gene	87	205	477
291	3	807	559	emb 283335 SP28	S. pneumoniae dxB, cap1A, B, C, D, E, F, G, H, I, J, K genes, dTDP-rhamnose biosynthesis genes and allA gene	90	170	249
291	4	1374	1099	gb M36180	Streptococcus pneumoniae transposase, (comA and comB) and SAICAR synthetase (putC) genes, complete cds	85	264	276

## TABLE 1

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	percent ident	HSP nt length	ORF nt length
293	1	3	1673	[emb Z67740 SPGY	S.pneumoniae gyrB gene and unknown orf	98	553	1671
296	1	1434	151	[emb Z47210 SPOE	S.pneumoniae dexB, capIA, capJB and capIC genes and orfs	99	430	1284
317	1	157	510	[emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	89	353	354
325	2	1237	485	[emb Z83335 SPZ8	S.pneumoniae dexB, capIIA,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	91	299	753
326	1	1	462	[emb Z82001 SPZ8	S.pneumoniae pcPA gene and open reading frames	100	233	462
327	1	603	64	[emb Z83335 SPZ8	S.pneumoniae dexB, capIIA,B,C,D,E,F,G,H,I,J,K) genes, qTOP-rhamnose biosynthesis genes and alia gene	94	89	540
334	1	153	545	[gb U41735]	Streptococcus pneumoniae peptide methionine sulfoxide reductase (msrA) and homoserine kinase homolog (thrB) genes, complete cds	87	91	393
336	1	308	93	[emb Z26850 SPAT	S.pneumoniae (M222) genes for ATPase a subunit, ATPase b subunit and ATPase c subunit	97	102	216
360	1	1	519	[emb Z67739 SPPA	S.pneumoniae parC, parE and transposase genes and unknown orf	95	435	519
360	4	1598	1960	[emb Z83335 SPZ8	S.pneumoniae dexB, capIIA,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	94	353	363
362	1	673	2	[emb Z83335 SPZ8	S.pneumoniae dexB, capIIA,B,C,D,E,F,G,H,I,J,K) genes, dTDP-rhamnose biosynthesis genes and alia gene	95	63	672
362	2	1168	728	[gb U04047]	Streptococcus pneumoniae SS2 dextran glucosidase gene and insertion sequence IS1202 transposase gene, Complete cds	96	441	441
384	1	347	111	[emb X85787 SPCP	S.pneumoniae dexB, cpsI4A, cpsI4B, cpsI4C, cpsI4D, cpsI4E, cpsI4G, cpsI4H, cpsI4J, cpsI4K, cpsI4L, tsaA genes	94	54	237

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
228	2	1760	1942	pir F60663 F606	translation elongation factor Tu Streptococcus oralis	100	100	183
319	1	2	205	gi 984927	neomycin phosphotransferase (cloning vector pBSL99)	100	100	204
260	1	2	1138	pir F60663 F606	translation elongation factor Tu Streptococcus oralis	99	98	1137
25	2	486	1394	gi 1574895	hypothetical (Haemophilus influenzae)	98	96	909
94	2	685	1002	gi 310627	phosphoenolpyruvate:sugar phosphotransferase system HPr Streptococcus mutans	98	93	318
312	1	190	2	gi 347999	ATP-dependent protease proteolytic subunit Streptococcus salivarius	98	95	189
329	1	1	807	gi 928848	inosine monophosphate dehydrogenase Streptococcus pyogenes	98	94	807
336	2	290	589	gi 987030	lacZ gene product (unidentified cloning vector)	98	98	300
181	9	5948	7166	gi 153755	phospho-beta-D-galactosidase (EC 3.2.1.85) Lactococcus lactis cremoris	97	94	1419
312	2	1044	361	gi 347998	urecili phosphoribosyltransferase Streptococcus salivarius	97	88	684
32	8	6575	7486	sp P7214 ERA_S	GTP-BINDING PROTEIN ERA HOMOLOG	96	91	912
94	3	951	2741	gi 153615	phosphoenolpyruvate:sugar phosphotransferase system enzyme I Streptococcus salivarius	96	92	1791
127	1	1	168	gi 581299	initiation factor IF-1 Lactococcus lactis	96	89	168
128	14	10438	11154	gi 1276873	Deob Streptococcus thermophilus	96	93	717
181	4	1362	1598	gi 46606	lacD polypeptide (AA 1-326) Staphylococcus aureus	96	80	237
218	1	1	834	gi 1743856	intragenetic coaggregation-relevant adhesin Streptococcus gordonii	96	93	834
319	2	115	441	gi 208225	heat-shock protein 82/neomycin phosphotransferase fusion protein (hsp82-neo) (unidentified cloning vector)	96	96	327
54	12	8622	10967	gnl prijd100972	Pyruvate formate-lyase Streptococcus mutans	95	89	2346
181	2	606	1289	gi 149396	lacD Lactococcus lactis	95	89	684
46	3	3410	3045	gi 1850606	YlxM Streptococcus mutans	94	86	366
89	10	7972	7337	gi 703442	chymidine kinase Streptococcus gordonii	94	86	636
148	9	6431	7154	gi 995767	UDP-glucose pyrophosphorylase Streptococcus pyogenes	94	85	924
160	7	4430	5848	gi 153573	H+ ATPase Enterococcus faecalis	94	87	1419
2	3	4598	3513	gi 153763	plasma receptor Streptococcus pyogenes	93	86	1086
12	8	7877	6204	gi 1103865	formyl-tetrahydrofolate synthetase Streptococcus mutans	93	84	1674

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
65	11	4734	5120	gi 40150	L14 protein (AA 1-122) [Bacillus subtilis]	93	87	387
68	1	53	1297	gi 47341	antitumor protein [Streptococcus pyogenes]	93	87	1245
80	1	3	299	gnl PID d101166	ribosomal protein S7 [Bacillus subtilis]	93	84	297
127	3	695	1093	gi 142462	ribosomal protein S11 [Bacillus subtilis]	93	86	399
160	5	1924	3462	gi 1773264	ATPase, alpha subunit [Streptococcus mutans]	93	85	1539
211	5	1757	3047	gi 535273	aminopeptidase C [Streptococcus thermophilus]	93	82	711
262	1	16	564	gi 149394	lacB [Lactococcus lactis]	93	90	549
366	1	197	3	gi 295259	cryptophan synthase beta subunit [Synecocystis sp.]	93	91	195
25	3	1392	1976	gi 1574496	hypothetical [Haemophilus influenzae]	92	80	585
36	21	2081	19927	gi 310632	hydrophobic membrane protein [Streptococcus gordonii]	92	86	855
181	3	1265	1534	gi 149196	lacD [Lactococcus lactis]	92	83	270
181	7	1662	4060	gi 149410	enzyme III [Lactococcus lactis]	92	83	399
32	4	5631	3937	gnl PID 294090	fibronectin-binding protein-like protein A [Streptococcus gordonii]	91	85	1695
46	2	1054	1462	gi 1850607	signal recognition particle Fth [Streptococcus mutans]	91	84	1593
65	10	4442	4726	pir SI7865 SI78	ribosomal protein S17 - Bacillus stearothermophilus	91	80	285
77	2	260	1900	gi 287871	groEL gene product [Lactococcus lactis]	91	82	1641
84	1	2	2056	gi 871784	Clp-like ATP-dependent protease binding subunit [Bos taurus]	91	79	2055
99	8	10750	9272	gi 153740	sucrose phosphorylase [Streptococcus mutans]	91	84	1479
99	9	11947	111072	gi 151739	membrane protein [Streptococcus mutans]	91	78	876
127	5	2065	2469	pir S07223 R585	ribosomal protein L17 - Bacillus stearothermophilus	91	78	405
132	6	9539	9390	gi 143065	hubst [Bacillus stearothermophilus]	91	89	150
137	8	4765	6153	gnl PID d100347	Na <sup>+</sup> ATPase beta subunit [Enterococcus hirae]	91	79	1389
151	7	11119	9734	gi 1815614	glutamine synthetase type I [Streptococcus agalactiae]	91	82	1386
201	2	1798	278	gi 2208998	dextran glucosidase DexS [Streptococcus suis]	91	79	1521
222	2	673	1839	gi 153741	ATP-binding protein [Streptococcus mutans]	91	85	1167
293	5	4113	4400	gi 1196921	unknown protein [insertion sequence IS861]	91	71	288
32	7	6166	6570	pir A36933 A369	diacylglycerol kinase homolog - Streptococcus mutans	90	77	405

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	2	841	527	[gi1196921]	unknown protein (insertion sequence IS861)	90	70	315
48	27	120908	19757	[gnl1P10]e274705	lactate oxidase (Streptococcus iniae)	90	80	1152
55	21	119777	18515	[gnl1P10]e221213	ClpX protein (Bacillus subtilis)	90	75	1263
56	2	717	977	[gi11710133]	flagellar filament cap (Burrelia burgdorferi)	90	50	261
65	1	1	606	[gi1165303]	L3 (Bacillus subtilis)	90	75	606
114	1	2	988	[gi1153562]	aspartate beta-semialdehyde dehydrogenase (EC 1.2.1.11) (Streptococcus mutans)	90	80	987
120	1	1345	827	[gi1407880]	ORF1 (Streptococcus equisimilis)	90	75	519
159	12	7690	8298	[gi1143012]	GMP synthetase (Bacillus subtilis)	90	84	609
166	4	4076	3282	[gi11661179]	high affinity branched chain amino acid transport protein (Streptococcus mutans)	90	78	795
183	1	28	1395	[gi1308858]	ATP:pyruvate 2-O-phosphotransferase (Lactococcus lactis)	90	76	1368
191	3	2891	1662	[gi1149521]	cryptophan synthase beta subunit (Lactococcus lactis)	90	78	1230
198	2	1551	436	[gi12323342]	(AF014460) CcpA (Streptococcus mutans)	90	76	1116
305	1	37	783	[gi11573551]	asparagine synthetase A (asnA) (Haemophilus influenzae)	90	80	747
8	3	2285	3343	[gi1149434]	putative (Lactococcus lactis)	89	78	1059
46	8	7577	7362	[pir1A4543]A454	ribosomal protein L19 - Bacillus stearothermophilus	89	76	216
49	9	8363	10342	[gi1153792]	recp peptide (Streptococcus pneumoniae)	89	83	1980
51	14	18410	19447	[gi1308857]	ATP-D-fructose 6-phosphate 1-phosphotransferase (Lactococcus lactis)	89	81	1018
57	11	9686	10669	[gnl1P10]d100932	H2O-forming NADH Oxidase (Streptococcus mutans)	89	77	984
65	5	2418	2786	[gi1165307]	S19 (Bacillus subtilis)	89	81	369
65	8	3806	4225	[sp1P14577]RL16	50S RIBOSOMAL PROTEIN L16	89	82	420
65	18	8219	8719	[gi1143417]	ribosomal protein S5 (Bacillus stearothermophilus)	89	76	501
73	9	6337	5315	[gi1532204]	prs (Listeria monocytogenes)	89	70	1023
76	3	3360	1465	[gnl1P10]e200671	lepA gene product (Bacillus subtilis)	89	76	1896
99	10	12818	11919	[gi1153738]	membrane protein (Streptococcus mutans)	89	73	900
120	2	1552	1300	[gi1407881]	stringent response-like protein (Streptococcus equisimilis)	89	79	2253
122	5	4512	2791	[gnl1P10]e280490	unknown (Streptococcus pneumoniae)	89	81	1722

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
176	1	669	4	gi 47394	5-oxoprolyl-peptidase (Streptococcus pyogenes)	89	78	666
177	6	3050	3934	gi 912423	putative (Lactococcus lactis)	89	71	885
181	8	4033	5751	gi 149411	enzyme III (Lactococcus lactis)	89	80	1719
211	4	3149	2793	gi 535273	aminopeptidase C (Streptococcus thermophilus)	89	83	357
361	1	431	838	gi 1196922	unknown protein (insertion sequence IS861)	89	70	408
34	17	11839	10535	sp P30053 SVH_S	HISTIDYL-TRNA SYNTHETASE (EC 6.1.1.21) (HISTIDINE--TRNA LIGASE) (HISRS)	88	78	1305
38	3	1646	2623	gi 2058544	putative ABC transporter subunit ComYA (Streptococcus gordonii)	88	78	978
54	1	3	227	gnl P10 dl01370	Yqgu (Bacillus subtilis)	88	66	225
57	2	611	1468	gnl P10 el34943	putative reductase I (Saccharomyces cerevisiae)	88	75	858
65	13	5497	6069	pir A29102 K58S	ribosomal protein L5 - Bacillus stearothermophilus	88	75	573
65	20	9030	9500	gi 2078381	ribosomal protein L15 (Staphylococcus aureus)	88	83	471
78	3	3636	1108	gnl P10 dl00781	lysyl-aminopeptidase (Lactococcus lactis)	88	80	2529
106	12	12965	12054	gi 2407215	(AF017421) putative heat shock protein HtpK (Streptococcus gordonii)	88	72	912
107	2	219	962	gnl P10 el33862	putative acylneuraminate lyase (Clostridium tertium)	88	75	744
111	8	14073	10420	gi 402363	RNA polymerase beta-subunit (Bacillus subtilis)	88	74	3654
126	9	13096	12062	gnl P10 el311468	unknown (Bacillus subtilis)	88	74	1035
140	17	19143	18874	gi 1573659	H. influenzae predicted coding region HI0659 (Haemophilus influenzae)	88	61	270
144	1	394	555	gnl P10 el274705	lactate oxidase (Streptococcus faecalis)	88	75	162
148	4	2723	3493	gi 1591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	88	68	771
160	8	5853	6278	gi 1773267	ATPase, epsilon subunit (Streptococcus mutans)	88	65	426
177	4	1770	2885	gi 149426	putative (Lactococcus lactis)	88	72	1116
211	6	4140	3613	gi 1535273	aminopeptidase C (Streptococcus thermophilus)	88	74	528
231	4	580	957	gi 40186	homologous to E. coli ribosomal protein L27 (Bacillus subtilis)	88	70	378
260	5	2387	2998	gi 1196922	unknown protein (insertion sequence IS861)	88	69	612
291	6	2017	3375	gnl P10 dl00571	adenylosuccinate synthetase (Bacillus subtilis)	88	75	1359
319	4	658	317	gi 603578	serine/threonine kinase (Phycophthora capsici)	88	88	302
319	5	4353	4514	gi 153672	lactose repressor (Streptococcus mutans)	87	56	162



TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
49	10	10660	10929	gi 1196921	unknown protein (insertion sequence IS861)	87	72	270
65	7	3140	3808	gi 1165109	S3 (Bacillus subtilis)	87	73	669
65	15	6623	7039	gi 1044978	ribosomal protein S8 (Bacillus subtilis)	87	73	417
75	8	5411	6625	gi 1877422	galactokinase (Streptococcus mutans)	87	78	1215
80	2	703	2805	gnl PID d101166	elongation factor G (Bacillus subtilis)	87	76	2103
82	1	541	248	gi 1196921	unknown protein (insertion sequence IS861)	87	69	294
140	23	25033	23897	gnl PID e254999	phenylalanyl-tRNA synthetase beta subunit (Bacillus subtilis)	87	74	1137
214	14	10441	8516	gi 2281305	glucose inhibited division protein homolog GIDA (Lactococcus lactis cremoris)	87	75	1926
220	2	2742	874	gnl PID e324358	product highly similar to elongation factor EF-G (Bacillus subtilis)	87	73	1869
260	4	2096	2389	gi 1196921	unknown protein (insertion sequence IS861)	87	72	294
323	1	27	650	gi 897795	30S ribosomal protein (pediococcus acidilactici)	87	73	624
357	1	154	570	gi 1044978	ribosomal protein S8 (Bacillus subtilis)	87	73	417
49	11	10927	11445	gi 1196922	unknown protein (insertion sequence IS861)	86	63	519
59	12	7461	9224	gi 1951051	relaxase (Streptococcus pneumoniae)	86	68	1764
65	4	1553	2401	pit A02759 8585	ribosomal protein L2 - Bacillus stearothermophilus	86	77	849
65	23	10957	11610	gi 144074	adenylate kinase (Lactococcus lactis)	86	76	654
82	4	4374	4856	gi 153745	mannitol-specific enzyme III (Streptococcus mutans)	86	72	483
102	4	4270	4986	gnl PID e264705	OMP decarboxylase (Lactococcus lactis)	86	76	717
106	6	7824	6880	gnl PID e137598	aspartate transcarbamylase (Lactobacillus leichmannii)	86	68	945
107	1	1	273	gnl PID e339862	putative acylneuraminate lyase (Clostridium tertium)	86	71	273
111	7	10432	6710	gnl PID e228283	DNA-dependent RNA polymerase (Streptococcus pyogenes)	86	80	3723
131	9	5704	4892	gi 1661193	polipoprotein diacylglycerol transferase (Streptococcus mutans)	86	71	813
134	7	6430	7980	gi 2388637	glycerol kinase (Enterococcus faecalis)	86	73	1551
146	11	7473	6583	gi 1591731	acetylglutamate kinase (Methanococcus jannaschii)	86	72	891
153	2	595	2010	gi 2160707	dipeptidase (Lactococcus lactis)	86	78	1416
154	1	2	1435	gi 1857246	6-phosphogluconate dehydrogenase (Lactococcus lactis)	86	74	1434

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
161	5	5025	6284	gi 47529	Unknown (Streptococcus salivarius)	86	66	1260
184	1	2	1403	gi 642667	NADP-dependent glyceraldehyde-3-phosphate dehydrogenase (Streptococcus mutans)	86	73	1482
210	8	1659	6571	gi 151661	translational initiation factor IF2 (Enterococcus faecium)	86	76	2913
250	1	2	187	gi 1573551	asparagine synthetase A (asaA) (Haemophilus influenzae)	86	68	186
36	4	2644	3909	gi 2149909	cell division protein (Enterococcus faecalis)	85	73	1266
38	4	2475	3587	gi 2058545	putative ABC transporter subunit ComVB (Streptococcus gordonii)	85	72	1113
38	5	3577	3915	gi 2058546	ComYC (Streptococcus gordonii)	85	80	339
57	5	2797	3789	gnl PI0101316	YqfJ (Bacillus subtilis)	85	72	993
82	5	4915	6054	gi 153746	mannitol-phosphate dehydrogenase (Streptococcus mutans)	85	68	1140
83	15	14690	15793	gi 143371	phosphoribosyl aminoimidazole synthetase (PUR-M) (Bacillus subtilis)	85	69	1104
87	2	1417	2188	gi 1184967	ScrR (Streptococcus mutans)	85	69	972
108	3	2666	3154	gi 153566	ORF (19K protein) (Enterococcus faecalis)	85	67	489
127	2	312	692	gi 1044989	ribosomal protein S13 (Bacillus subtilis)	85	72	381
128	3	1534	2409	gi 1685110	tetrahydrofolate dehydrogenase/cyclohydrolase (Streptococcus thermophilus)	85	71	876
137	7	2962	4767	gnl PI0100347	Na <sup>+</sup> ATPase alpha subunit (Enterococcus hirae)	85	74	1806
170	2	2622	709	gnl PI0102006	(AB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI, H. INFLUENZAE AND NEISSERIA MENINGITIDIS. (Bacillus subtilis)	85	70	1914
187	5	3760	4386	gi 727436	putative 20-kDa protein (Lactococcus lactis)	85	65	627
233	2	728	1873	gi 1163116	ORF-5 (Streptococcus pneumoniae)	85	67	1146
234	3	962	1255	gi 2293155	(AF008220) Y1A (Bacillus subtilis)	85	61	294
240	1	309	1931	gi 143597	CTP synthetase (Bacillus subtilis)	85	70	1623
6	1	199	1521	gi 508979	GTP-binding protein (Bacillus subtilis)	84	72	1323
10	4	4375	3443	gnl PI0103982	putative acylneuraminate lyase (Clostridium tertium)	84	70	933
14	1	63	2093	gi 520753	DNA topoisomerase I (Bacillus subtilis)	84	69	2031
19	4	1793	2593	gi 2152484	(AF005098) RNASEH II (Lactococcus lactis)	84	68	801
20	17	17720	19687	gnl PI0100584	cell division protein (Bacillus subtilis)	84	71	1968
22	28	21723	20884	gi 299163	alanine dehydrogenase (Bacillus subtilis)	84	68	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
10	10	7730	6792	[gnl pid d100296]	[fructokinase (Streptococcus mutans)]	84	75	939
33	9	5650	5300	[gi 147194]	[phnA protein (Escherichia coli)]	84	71	351
36	22	121531	20772	[gi 310631]	[ATP binding protein (Streptococcus gordonii)]	84	72	780
48	4	2837	2505	[gi 882609]	[6-phospho-beta-glucosidase (Escherichia coli)]	84	69	333
58	1	41	1516	[gi 450849]	[amylase (Streptococcus bovis)]	84	73	1476
59	10	6715	7116	[gi 951053]	[ORF10, putative (Streptococcus pneumoniae)]	84	74	402
62	1	21	644	[gi 806487]	[ORF211; putative (Lactococcus lactis)]	84	66	624
65	17	7779	8207	[gi 1044980]	[ribosomal protein L18 (Bacillus subtilis)]	84	73	429
65	21	9507	10397	[gi 44073]	[SecY protein (Lactococcus lactis)]	84	68	891
106	4	5474	2262	[gnl pid d109387]	[carbamoyl-phosphate synthase (Lactobacillus plantarum)]	84	73	3213
159	1	147	4	[gi 806487]	[ORF211; putative (Lactococcus lactis)]	84	63	144
163	4	4690	5910	[gi 2293164]	[AF008220] SAM synthase (Bacillus subtilis)]	84	69	1221
192	1	46	1308	[gi 495046]	[tripeptidase (Lactococcus lactis)]	84	73	1263
348	1	671	6	[gi 1787753]	[AE000245] (146; 79 pct identical to 336 amino acids of ADH1 2YMO SW: P20368 but has 10 additional N-ter residues (Escherichia coli)]	84	71	666
3	4	1572	3575	[gi 143766]	[(thrSv) (EC 6.1.1.3) (Bacillus subtilis)]	83	65	2004
9	6	3893	3417	[gnl pid d100576]	[single strand DNA binding protein (Bacillus subtilis)]	83	68	477
17	15	7426	8457	[gi 520738]	[comA protein (Streptococcus pneumoniae)]	83	66	1032
20	12	13860	14144	[gnl pid d100583]	[unknown (Bacillus subtilis)]	83	61	285
23	4	3358	2606	[gi 1788294]	[AE000290] (238; This 238 aa orf is 40 pct identical (5 gaps) to 231 residues of an approx. 248 aa protein Y8AC_ECOLI SW: P24237 (Escherichia coli)]	83	74	753
28	6	3304	3005	[gi 1573659]	[H. influenzae predicted coding region H10659 (Haemophilus influenzae)]	83	57	300
35	7	5108	3867	[gi 311707]	[hypothetical nucleotide binding protein (Acholeplasma laidlawii)]	83	63	1242
55	19	17932	17528	[gi 533085]	[ORF_114] (Escherichia coli)]	83	59	405
55	20	18539	17919	[gi 496558]	[orfX (Bacillus subtilis)]	83	69	621
65	6	2795	3142	[gi 1165308]	[L22 (Bacillus subtilis)]	83	64	348
68	6	6877	6683	[gi 1213494]	[immunoglobulin A1 protease (Streptococcus pneumoniae)]	83	54	195

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
87	15	15112	14771	gnl pid d133522	putative rpo2 protein (Bacillus subtilis)	83	54	342
96	12	8963	9631	gi 47394	[5-oxopropyl]-peptidase (Streptococcus pyogenes)	83	73	669
98	1	3	263	gi 1183885	glutamine-binding subunit (Bacillus subtilis)	83	55	261
120	4	7170	5233	gi 110630	zinc metalloprotease (Streptococcus gordonii)	83	72	1938
127	7	2998	4347	gi 1500567	M. Jannaschii predicted coding region (J1665 (Methanococcus jannaschii))	83	72	1350
137	1	3	440	gi 472918	[v-type Na-ATPase (Enterococcus hirae)]	83	60	438
160	6	3466	4356	gi 1773265	ATPase, gamma subunit (Streptococcus mutans)	83	67	891
214	4	2278	2964	gi 663279	transposase (Streptococcus pneumoniae)	83	72	687
226	3	2367	2020	gi 142154	thioredoxin (Synecococcus PCC6301)	83	58	348
301	1	3	1049	gi 40046	phosphoglucose isomerase A (AA 1-469) (Bacillus stearothermophilus)	83	67	1047
303	2	1155	1931	gi 289282	glutamy-tRNA synthetase (Bacillus subtilis)	83	67	777
6	17	15370	14318	gi 633147	ribose-phosphate pyrophosphokinase (Bacillus caldolyticus)	82	64	1053
7	1	299	96	gi 143648	ribosomal protein L28 (Bacillus subtilis)	82	69	204
9	3	1479	1090	gi 385178	unknown (Bacillus subtilis)	82	46	390
9	7	4213	3899	gnl pid d100576	ribosomal protein S6 (Bacillus subtilis)	82	60	315
12	6	4688	3942	gnl pid d100571	unknown (Bacillus subtilis)	82	68	747
22	17	13422	114837	gi 520754	putative (Bacillus subtilis)	82	69	1416
22	18	14897	15658	gnl pid d101929	uridine monophosphate kinase (Synecocystis sp.)	82	62	762
33	16	11471	10641	gnl pid d101190	ORF4 (Streptococcus mutans)	82	68	831
35	9	7400	6255	gi 1881543	[UDP-N-acetylglucosamine-2-epimerase (Streptococcus pneumoniae)]	82	68	1146
40	10	8003	7533	gi 1173519	riboflavin synthase beta subunit (Actinobacillus pleuropneumoniae)	82	68	471
48	32	23159	23437	gi 1930092	outer membrane protein (Campylobacter jejuni)	82	61	279
52	14	13833	14765	gi 142521	[deoxyribodipyrimidine photolyase (Bacillus subtilis)]	82	61	933
60	4	4737	1849	gnl pid d101221	[AB001810] uvrA (Deinococcus radiodurans)	82	66	2889
62	4	2131	1457	gi 2246749	[AF009622] thioredoxin reductase (Listeria monocytogenes)	82	63	675
71	11	18586	17518	gnl pid d132206	[as-1.4-galactosyltransferase (Streptococcus pneumoniae)]	82	60	933
73	13	9222	7837	gnl pid d100586	unknown (Bacillus subtilis)	82	65	1386

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	1	1	3771	gnl p10 d101199	alkaline amylopullulanase [Bacillus sp.]	82	68	3771
83	9	3696	3983	gnl p10 e305362	unnamed protein product [Streptococcus thermophilus]	82	52	288
86	11	110776	9194	gi 683583	5-enolpyruvylshikimate-3-phosphate synthase [Lactococcus lactis]	82	67	1383
89	12	8295	9752	gi 40025	homologous to E. coli 50K [Bacillus subtilis]	82	66	1458
115	9	10147	8812	gnl p10 d102090	[AB003927] phospho-beta-galactosidase 1 [Lactobacillus gasseri]	82	74	1536
118	1	1	1332	gnl p10 d100579	[seryl-tRNA synthetase [Bacillus subtilis]	82	71	1332
151	1	4657	6246	pir S06097 S060	type 1 site-specific deoxyribonuclease (EC 3.1.21.3) CtrA chain S - Citrobacter freundii	82	66	1590
173	6	4183	3503	gi 2313836	[AE000584] conserved hypothetical protein [Helicobacter pylori]	82	68	681
177	12	5481	7442	gnl p10 d101999	[AB001341] NcrB [Escherichia coli]	82	58	1962
193	2	178	576	pir S08564 RJB5	ribosomal protein S9 - Bacillus stearothermophilus	82	70	399
245	2	258	845	gi 146402	EcoA type 1 restriction-modification enzyme S subunit [Escherichia coli]	82	68	588
9	5	3400	3146	gnl p10 d100576	ribosomal protein S18 [Bacillus subtilis]	81	66	255
16	7	7484	8413	gi 1100074	[cryptophenyl-tRNA synthetase [Clostridium longisporum]	81	70	930
20	11	10308	13820	gnl p10 d100583	transcription-repair coupling factor [Bacillus subtilis]	81	63	3513
38	2	1232	1606	gi 2058543	putative DNA binding protein [Streptococcus gordonii]	81	63	375
45	2	3081	1751	gi 460259	enolase [Bacillus subtilis]	81	67	1311
46	1	2	1267	gi 431231	uracil permease [Bacillus caldolyticus]	81	61	1266
48	3	2453	1440	gnl p10 d100453	Mannosephosphate isomerase [Streptococcus mutans]	81	70	1014
54	2	1106	336	gi 154752	transport protein [Agrobacterium tumefaciens]	81	64	771
65	22	110306	10821	gi 44073	[SecY protein [Lactococcus lactis]	81	66	516
89	4	3874	2603	gi 556886	serine hydromethyltransferase [Bacillus subtilis]	81	69	1272
99	16	119126	18929	gi 2313526	[AE000557] H. pylori predicted coding region HP0411 [Helicobacter pylori]	81	75	198
106	7	8373	7822	gnl p10 e199384	[pyrK [Lactobacillus plantarum]	81	61	552
108	6	5054	6877	gi 1669939	[group B oligopeptidase PepB [Streptococcus agalactiae]	81	66	1824
113	15	115899	18283	pir S09411 S094	[spolIIE protein - Bacillus subtilis	81	65	2385
128	5	3359	1634	gi 1685111	[orf1091 [Streptococcus thermophilus]	81	69	276

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	CRF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	1	830	3211	gi 304896	EcoE type I restriction-modification enzyme R subunit [Escherichia coli]	81	59	2382
159	11	6722	7837	gi 2239288	GMP synthetase [Bacillus subtilis]	81	69	1116
170	1	739	458	gnl PID102006	FUNCTION UNKNOWN. [Bacillus subtilis]	81	55	282
191	2	1759	893	gi 149522	cryptophan synthase alpha subunit [Lactococcus lactis]	81	65	867
214	3	2290	1594	gi 157587	reverse transcriptase endonuclease [Drosophila virilis]	81	43	297
217	4	4415	4008	gi 466473	cellobiose phosphotransferase enzyme II' [Bacillus stearothermophilus]	81	59	408
262	2	569	868	gi 153675	Lagatose 6-P kinase [Streptococcus mutans]	81	68	300
299	1	663	4	gnl PID1030154	StySKI methylase [Salmonella enterica]	81	60	660
366	2	376	83	gi 149521	cryptophan synthase beta subunit [Lactococcus lactis]	81	65	294
32	10	8766	9242	gi 1216490	DNA/pantothenate metabolism flavoprotein [Streptococcus mutans]	80	64	477
17	11	6050	5748	gnl PID10305362	unnamed protein product [Streptococcus thermophilus]	80	67	303
17	16	8455	9066	gi 703126	leucocin A translocator [Leuconostoc gelidium]	80	59	612
18	1	2440	1613	gi 1591672	phosphate transport system ATP-binding protein [Methanococcus jannaschii]	80	58	828
27	3	4248	1579	gi 452309	valyl-tRNA synthetase [Bacillus subtilis]	80	69	2670
28	7	3671	1288	gi 1573660	H. influenzae predicted coding region H10660 [Haemophilus influenzae]	80	63	384
32	2	902	1933	gnl PID10264499	dihydroorotate dehydrogenase B [Lactococcus lactis]	80	66	1032
19	1	1	1266	gnl PID10234078	hom [Lactococcus lactis]	80	63	1266
52	5	4363	3593	gi 1181884	ATP-binding subunit [Bacillus subtilis]	80	57	771
54	5	4550	4744	gi 2198820	[AF004225] Cux/CDP(181); Cux/COP homeoprotein [Mus musculus]	80	60	195
59	11	7109	7486	gi 951052	ORF9, putative [Streptococcus pneumoniae]	80	68	378
65	3	1230	1550	pir A02815 RS8S	[ribosomal] protein L23 - Bacillus stearothermophilus	80	69	321
65	12	5174	5503	pir A02819 RS8S	[ribosomal] protein L24 - Bacillus stearothermophilus	80	70	330
66	9	9884	10687	gi 2311836	[AE000584] conserved hypothetical protein [Helicobacter pylori]	80	66	804
82	2	648	2438	gi 1622991	mannotol transport protein [Bacillus stearothermophilus]	80	65	1791
85	1	950	630	gi 528995	polyketide synthase [Bacillus subtilis]	80	46	321
89	8	6810	5779	gi 1853776	peptide chain release factor 1 [Bacillus subtilis]	80	63	1092
93	12	8718	7438	gnl PID101959	hypothetical protein [Synechocystis sp.]	80	60	1281

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	5	6854	5751	[gnl PID e199186]	glutaminase of carboxyl-phosphate synthase (Lactobacillus plantarum)	80	65	1104
109	2	2160	1450	[gi 40036]	phoP gene product (Bacillus subtilis)	80	59	711
124	9	4246	3953	[gnl PID d102254]	30S ribosomal protein S16 (Bacillus subtilis)	80	65	294
128	8	5148	6428	[gi 2281308]	phosphoenolpyruvate carboxykinase (Lactococcus lactis cremoris)	80	66	1281
137	19	12665	11376	[gi 159109]	NADP-dependent glutamate dehydrogenase (Giardia intestinalis)	80	68	1290
140	19	119699	119457	[gi 517210]	putative transposase (Streptococcus pyogenes)	80	70	243
158	2	2474	984	[gi 18877423]	galactose-1-P-uridylyl transferase (Streptococcus mutans)	80	65	1491
171	10	7474	7728	[gi 197800]	cyclophilin C-associated protein (Mus musculus)	80	60	255
181	1	2	619	[gi 149395]	lacC (Lactococcus lactis)	80	66	618
313	1	27	539	[gi 143467]	ribosomal protein S4 (Bacillus subtilis)	80	70	513
329	2	1652	858	[gi 533080]	RecF protein (Streptococcus pyogenes)	80	63	795
331	1	2	958	[gi 442360]	ClpC adenosine triphosphatase (Bacillus subtilis)	80	58	957
8	7	4312	5580	[gi 149435]	putative (Lactococcus lactis)	79	64	1269
23	1	1175	135	[gi 1542975]	AcbB (Thermotoga bacterium thermophilum)	79	61	1041
33	14	9244	8201	[gnl PID e253891]	UDP-glucose 4-epimerase (Bacillus subtilis)	79	62	1044
36	3	1242	2633	[gnl PID e324238]	ftsA (Enterococcus hirae)	79	58	1392
38	13	7155	8378	[gi 405134]	acetate kinase (Bacillus subtilis)	79	58	1224
55	7	9011	8229	[gi 1146234]	dihydrodipicolinate reductase (Bacillus subtilis)	79	56	783
65	19	8661	8915	[gi 2078380]	ribosomal protein L30 (Staphylococcus aureus)	79	68	255
69	4	3678	2128	[gnl PID e311452]	unknown (Bacillus subtilis)	79	64	1551
69	9	7881	7279	[gi 677850]	hypothetical protein (Staphylococcus aureus)	79	59	603
72	10	8491	9783	[gnl PID d101091]	hypothetical protein (Synecocystis sp.)	79	62	1293
80	3	2906	7300	[gi 143342]	polymerase III (Bacillus subtilis)	79	65	4395
82	14	13326	15689	[gnl PID e25509]	hypothetical protein (Bacillus subtilis)	79	65	2364
86	13	12233	11118	[gi 683582]	prephenate dehydrogenase (Lactococcus lactis)	79	58	1116
92	3	940	1314	[gi 533286]	triosephosphate isomerase (Lactococcus lactis)	79	65	795
98	6	4023	4742	[gnl PID d100762]	LivG protein (Salmonella typhimurium)	79	63	720

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
99	12	16315	14150	gi 153736	[a-galactosidase (Streptococcus mutans)]	79	64	2166
107	7	5684	6406	gi 460080	[D-alanine:D-alanine ligase-related protein (Enterococcus faecalis)]	79	58	723
113	9	6858	8303	gi 466882	[ppa1; B1496_C2.189 (Mycobacterium leprae)]	79	64	1446
151	10	13424	12213	gi 450686	[3-phosphoglycerate kinase (Thermotoga maritima)]	79	60	1212
162	2	1158	3017	gi 506700	[CapD (Staphylococcus aureus)]	79	67	1860
177	5	2876	3052	gi 912423	[putative (Lactococcus lactis)]	79	61	177
177	8	4198	4561	gi 149429	[putative (Lactococcus lactis)]	79	61	366
187	3	2728	2907	gnl PID d102002	[AB001488] FUNCTION UNKNOWN. (Bacillus subtilis)]	79	53	180
189	7	3589	4350	gnl PIN e183449	[putative ATP-binding protein of ABC-type (Bacillus subtilis)]	79	61	762
191	5	4249	3449	gi 149519	[indoleglycerol phosphate synthase (Lactococcus lactis)]	79	66	801
211	3	1805	2737	gi 147404	[mannose permease subunit II-M-Hen (Escherichia coli)]	79	57	933
212	3	3863	3621	gnl PID e209004	[glutaredoxin-like protein (Lactococcus lactis)]	79	58	243
215	1	987	715	gi 2293242	[AF008220] arginine succinate synthase (Bacillus subtilis)]	79	64	273
323	2	510	781	gi 897795	[30S ribosomal protein (Pedicoccus acidilactici)]	79	67	252
380	1	694	2	gi 1184680	[polynucleotide phosphorylase (Bacillus subtilis)]	79	64	693
384	2	655	239	gi 143328	[phop protein (put.); putative (Bacillus subtilis)]	79	59	417
6	3	2820	4091	gi 853767	[UDP-N-acetylglucosamine 1-carboxyvinyltransferase (Bacillus subtilis)]	78	62	1272
8	1	50	1786	gi 149432	[putative (Lactococcus lactis)]	78	63	1737
9	1	351	124	gi 897793	[y98 gene product (Pedicoccus acidilactici)]	78	59	228
15	8	7364	8314	gnl PID d100585	[cysteine synthetase A (Bacillus subtilis)]	78	63	951
20	10	9738	10310	gnl PID d100581	[stage V sporulation (Bacillus subtilis)]	78	58	573
20	16	17165	17713	gi 49105	[hypoxanthine phosphoribosyltransferase (Lactococcus lactis)]	78	59	549
22	22	17388	18416	gnl PID d101315	[yqfE (Bacillus subtilis)]	78	60	1029
22	27	20971	20612	gi 299163	[alanine dehydrogenase (Bacillus subtilis)]	78	59	360
34	8	7407	7105	gi 41015	[aspartate-tRNA ligase (Escherichia coli)]	78	55	303
35	8	6257	5196	gi 1657644	[Cap8E (Staphylococcus aureus)]	78	60	1062



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
40	11	9287	8001	gi11173518	GTP cyclohydrolase II/ J 4-dihydroxy 2-butanone-4-phosphate synthase [Actinobacillus pleuropneumoniae]	78	58	1287
48	31	22422	23183	gi12314330	[AE000623] glutamine ABC transporter, ATP-binding protein (glnQ) [Helicobacter pylori]	78	58	762
52	2	2101	1430	gi11183887	integral membrane protein [Bacillus subtilis]	78	54	672
55	14	13605	12712	gnl PID102026	[AB002150] YbbP [Bacillus subtilis]	78	58	894
55	17	16637	15612	gnl PIDe313027	hypothetical protein [Bacillus subtilis]	78	51	1026
71	14	19756	19598	gi1179764	calcium channel alpha-1D subunit (Homo sapiens)	78	57	159
74	11	15031	14018	gi11573279	Holliday junction DNA helicase (ruvB) [Haemophilus influenzae]	78	57	1014
75	9	6623	7972	gi11877423	galactose-1-P-uridylyl transferase [Streptococcus mutans]	78	62	1350
81	12	12125	11306	gi11573607	beta-fucose isomerase (fucI) [Haemophilus influenzae]	78	66	1782
82	3	2423	4417	gi1153744	[ORF X; putative [Streptococcus mutans]	78	64	1995
83	18	16926	18500	gi1143373	[phosphoribosyl aminimidazole carboxy formyl formyltransferase/inosine monophosphate cyclohydrolase (pur-HIJ)] [Bacillus subtilis]	78	63	1575
83	20	20212	20775	gi1143364	[phosphoribosyl aminimidazole carboxylase I (pur-E) [Bacillus subtilis]	78	64	564
92	2	165	878	gnl PID101190	[ORF2 [Streptococcus mutans]	78	62	714
98	8	5863	6909	gi12331287	[AF013188] release factor 2 [Bacillus subtilis]	78	63	1047
113	3	1071	2741	gi1580914	[dnaX [Bacillus subtilis]	78	64	1671
127	4	1133	2071	gi1142463	[RNA polymerase alpha-core subunit [Bacillus subtilis]	78	59	939
132	1	2782	497	gi11561763	[pullulanase [Bacteroides thetaiotaomicron]	78	58	2286
135	4	2698	3537	gi11788036	[AE000269] NH3-dependent NAD synthetase [Escherichia coli]	78	66	840
140	24	26853	25823	gi11100077	[phospho-beta-glucosidase [Clostridium longisporum]	78	64	1431
150	5	4690	4514	gi1149464	[amino peptidase [Lactococcus lactis]	78	42	177
152	1	1	795	gi1639915	[NADH dehydrogenase subunit (Thunbergia alata]	78	43	795
182	4	4997	4110	gnl PIDe123528	[putative YhaP protein [Bacillus subtilis]	78	64	888
181	10	8651	7947	gi1149402	[lactose repressor (lacR; alt.) [Lactococcus lactis]	78	48	705
200	4	3627	4958	gnl PID100172	[invertase [Zymomonas mobilis]	78	61	1332
203	3	3230	3015	gi11174237	[Cyck [Pseudomonas fluorescens]	78	57	216

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
210	9	6789	7172	gi 580902	ORF6 gene product (Bacillus subtilis)	78	42	384
214	6	7810	2797	gnl PID d102049	P. haemolytica o-sialoglycoprotein endopeptidase; P36175 (660) transmembrane (Bacillus subtilis)	78	60	1014
214	13	6322	8163	gi 1377831	unknown (Bacillus subtilis)	78	62	1842
217	1	9	2717	gi 488430	alcohol dehydrogenase 2 (Entamoeba histolytica)	78	64	2709
222	3	2316	3098	gi 1573047	spore germination and vegetative growth protein (gerC2) (Haemophilus influenzae)	78	65	783
268	1	742	8	gi 517210	putative transposase (Streptococcus pyogenes)	78	65	735
276	1	223	753	gnl PID d100306	ribosomal protein L1 (Bacillus subtilis)	78	65	531
312	3	1567	1079	gi 289261	comE ORF2 (Bacillus subtilis)	78	54	489
319	1	117	794	gi 1916729	CadD (Staphylococcus aureus)	78	53	678
342	2	762	265	gi 1842419	phosphatidylglycerophosphate synthase (Bacillus subtilis)	78	59	498
383	1	737	3	gi 1184680	polynucleotide phosphorylase (Bacillus subtilis)	78	64	735
7	15	11921	11018	gi 1398855	carboxyltransferase beta subunit (Synchococcus PCC7942)	77	63	906
8	2	1698	2255	gi 149433	putative (Lactococcus lactis)	77	59	558
17	14	6948	7550	gi 520738	comA protein (Streptococcus pneumoniae)	77	60	603
30	12	9761	8967	gi 1000451	TreP (Bacillus subtilis)	77	43	795
36	14	11421	12131	gi 1573766	phosphoglyceromutase (gpmA) (Haemophilus influenzae)	77	64	711
55	3	1836	4096	gi 1708640	YeaB (Bacillus subtilis)	77	55	261
61	8	8377	8054	gi 1890649	multidrug resistance protein LmrA (Lactococcus lactis)	77	51	324
65	2	607	1254	gi 40103	ribosomal protein L4 (Bacillus stearothermophilus)	77	63	648
68	8	7509	7240	gi 47551	MRP (Streptococcus sulci)	77	68	270
69	1	1083	118	gnl PID e311493	unknown (Bacillus subtilis)	77	57	966
77	5	4583	4026	gnl PID e281578	hypothetical 12.2 kd protein (Bacillus subtilis)	77	60	558
83	14	13104	14552	gi 1590947	amidophosphoribosyltransferase (Methanococcus jannaschii)	77	56	1449
94	4	3006	5444	gnl PID e329895	(AJ000496) cyclic nucleotide-gated channel beta subunit (Rattus norvegicus)	77	66	2439
96	11	8518	8880	gi 551879	ORF 1 (Lactococcus lactis)	77	62	363
99	11	14082	12799	gi 153737	sugar-binding protein (Streptococcus mutans)	77	61	1284

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	2	361	1176	gi 148921	LicD protein (Haemophilus influenzae)	77	51	816
108	4	3152	4030	gi 1574730	cellulose resistance protein (cehB) (Haemophilus influenzae)	77	58	879
118	4	3520	3131	gi 1573900	D-alanine permease (dapA) (Haemophilus influenzae)	77	57	390
124	4	1796	1071	gi 1573162	tRNA (guanine-N1)-methyltransferase (trmD) (Haemophilus influenzae)	77	58	726
126	4	5909	4614	gnl pid d101163	Srb (Bacillus subtilis)	77	62	1296
128	2	630	1373	gnl pid d101328	Yqi2 (Bacillus subtilis)	77	58	744
130	1	1	1287	gnl pid e325013	hypothetical protein (Bacillus subtilis)	77	61	1287
139	5	4388	3639	gi 2293302	(AF008220) Yti4A (Bacillus subtilis)	77	59	750
140	11	10931	9582	gi 289284	cysteineyl-tRNA synthetase (Bacillus subtilis)	77	64	1350
140	18	19451	19263	gi 517210	putative transposase (Streptococcus pyogenes)	77	66	189
141	2	976	1683	gnl pid e157887	URF5 (aa 1-573) (Drosophila yakuba)	77	50	708
141	4	2735	5293	gi 556258	baeA (Listeria monocytogenes)	77	59	2539
144	2	671	2173	gnl pid d100585	lyyA1-tRNA rhythetase (Bacillus subtilis)	77	61	1503
163	5	6412	7398	gi 511015	dihydroorotate dehydrogenase A (Lactococcus lactis)	77	62	987
164	10	7841	7074	gnl pid d100964	homologue of iron dicitrate transport ATP-binding protein FecE of E. coli (Bacillus subtilis)	77	52	768
191	8	7257	5791	gi 149516	anthranilate synthase alpha subunit (Lactococcus lactis)	77	57	1467
198	8	5377	5177	gi 1573856	hypothetical (Haemophilus influenzae)	77	66	203
213	1	202	462	gi 1743860	Bica2 (Mus musculus)	77	50	261
250	2	231	509	gnl pid e334776	VibW protein (Bacillus subtilis)	77	60	279
289	3	1737	1276	gnl pid d100947	Ribosomal Protein L10 (Bacillus subtilis)	77	62	462
292	2	1399	668	gi 143004	transfer RNA-Gln synthetase (Bacillus stearothermophilus)	77	58	732
7	1	2714	1166	gnl pid d101824	peptide-chain-release factor 3 (Synchocystis sp.)	76	53	1569
7	23	18474	18235	gi 455157	acyl carrier protein (Cryptomonas phil)	76	57	240
9	8	5706	4362	gi 1146247	asparaginyl-tRNA synthetase (Bacillus subtilis)	76	61	1365
10	5	4531	4385	gnl pid e314495	hypothetical protein (Clostridium perfringens)	76	53	147
18	2	1615	842	gi 1591672	phosphate transport system ATP-binding protein (Methanococcus jannaschii)	76	56	774

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	37	2796	28173	gnl PID e13389	translation initiation factor IF1 (AA 1-172) [Bacillus stearothermophilus]	76	64	378
35	6	3869	2682	gi 1773346	Cap5G [Staphylococcus aureus]	76	61	1188
48	28	21113	21787	gi 2314328	[AE000623] glutamine ABC transporter, permease protein (glnP) [Helicobacter pylori]	76	52	675
52	12	12881	13786	gi 142521	deoxyribodipyrimidine photolyase [Bacillus subtilis]	76	58	906
55	10	11521	10571	gnl PID e283110	femD [Staphylococcus aureus]	76	61	951
57	8	7824	6559	gi 290561	ol88 [Escherichia coli]	76	47	1266
62	5	2406	2095	gnl PID e13024	hypothetical protein [Bacillus subtilis]	76	59	312
65	9	4223	4441	gi 40148	[L29 protein (AA 1-66) [Bacillus subtilis]	76	58	219
68	2	1328	2371	gnl PID e284233	anabolic ornithine carbamoyltransferase [Lactobacillus plantarum]	76	61	1044
69	8	7297	6005	gnl PID d101420	pyrimidine nucleoside phosphorylase [Bacillus stearothermophilus]	76	61	1293
73	12	7819	7267	gnl PID e243629	unknown [Mycobacterium tuberculosis]	76	53	573
74	5	8433	7039	gnl PID d102048	[C. thermocellum beta-glucosidase; P26208 1985] [Bacillus subtilis]	76	60	1395
80	5	7643	7916	gi 2314010	[AE000599] conserved hypothetical protein [Helicobacter pylori]	76	61	294
82	15	16019	16996	gi 1571900	[O-alanine permease (dagA) [Haemophilus influenzae]	76	56	978
83	19	18616	19884	gnl 143374	[phosphoribosyl glycineamide synthetase (PUR-D; gta start codon) [Bacillus subtilis]	76	60	1269
86	14	13409	12231	gi 143806	AroF [Bacillus subtilis]	76	58	1179
87	1	3	1442	gi 153804	sucrose-6-phosphate hydrolase [Streptococcus mutans]	76	59	1440
87	16	15754	15110	gnl PID e233500	putative Cmk protein [Bacillus subtilis]	76	56	645
93	4	1769	1539	gi 1574820	[1,4-alpha-glucan branching enzyme (glgB) [Haemophilus influenzae]	76	46	231
94	1	51	365	gi 144313	[6.0 kD ORF (Plasmid ColE1)]	76	73	315
116	2	2151	1678	gi 153841	[pneumococcal surface protein A [Streptococcus pneumoniae]	76	59	474
123	6	1442	5895	gi 1314297	ClpC ATPase [Listeria monocytogenes]	76	59	2454
126	2	2156	2932	gnl PID d101328	[Kq2 [Bacillus subtilis]	76	61	777
128	10	6973	7797	gi 944944	[purine nucleoside phosphorylase [Bacillus subtilis]	76	60	825
131	11	6186	5812	gi 1674310	[AE000058] Mycoplasma pneumoniae, MC085 homolog, from M genitalium [Mycoplasma pneumoniae]	76	47	375

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
139	4	3641	3192	[gi 2293302	[AF008220] YtqA [Bacillus subtilis]	76	53	450
140	14	14872	12536	[gi 1184680	[polynucleotide phosphorylase [Bacillus subtilis]	76	62	2337
143	2	2583	3905	[gi 143795	[transfer RNA-Tyr synthetase [Bacillus subtilis]	76	61	1323
170	6	5095	6114	[gnl p10 d100959	[ycgQ [Bacillus subtilis]	76	44	1020
180	2	1927	557	[gi 40019	[ORF 821 (aa 1-821) [Bacillus subtilis]	76	53	1371
191	7	5815	5228	[gi 551880	[anthranilate synthase beta subunit [Lactococcus lactis]	76	61	588
195	3	3829	2444	[gi 2149905	[D-glutamic acid adding enzyme [Enterococcus faecalis]	76	60	1386
200	3	1914	3629	[gi 431272	[lysis protein [Bacillus subtilis]	76	58	1716
201	1	431	207	[gi 2208998	[dextran glucosidase DaxS [Streptococcus suis]	76	57	225
214	2	1283	2380	[gi 663278	[transposase [Streptococcus pneumoniae]	76	55	1098
225	1	2338	3413	[gi 3552775	[ATP-binding protein [Escherichia coli]	76	56	1074
233	1	2	724	[gi 1163115	[neuraminidase B [Streptococcus pneumoniae]	76	60	723
347	1	523	38	[gi 537033	[ORF f356 [Escherichia coli]	76	60	486
356	2	842	165	[gi 2149905	[D-glutamic acid adding enzyme [Enterococcus faecalis]	76	61	678
366	3	734	348	[gi 149520	[phosphoribosyl anthranilate isomerase [Lactococcus lactis]	76	69	387
5	8	12599	11484	[gi 1574293	[flabrial transcription regulation repressor (pilB) [Haemophilus influenzae]	75	61	1116
6	13	12553	11894	[gnl p10 d102050	[ydhH [Bacillus subtilis]	75	51	660
9	10	7282	6062	[gi 142538	[aspartate aminotransferase [Bacillus sp.]	75	55	1221
10	12	8080	7940	[gi 149493	[SCRF1 methylase [Lactococcus lactis]	75	56	141
18	5	4266	3301	[gnl p10 d101319	[yqgH [Bacillus subtilis]	75	52	966
22	4	1838	2728	[gi 1373157	[orf-X; hypothetical protein; Method: conceptual translation supplied by author [Bacillus subtilis]	75	62	891
30	11	9015	7828	[gi 153801	[enzyme scr-11 [Streptococcus mutans]	75	64	1188
31	5	2362	2030	[gi 2293231	[AF008220] putative thioredoxin [Bacillus subtilis]	75	53	333
32	9	7484	8359	[gnl p10 d100560	[formamidopyrimidine-DNA glycosylase [Streptococcus mutans]	75	61	876
33	4	1715	1448	[gi 413976	[ipa 52r gene product [Bacillus subtilis]	75	53	288
33	10	6470	5769	[gi 533105	[unknown [Bacillus subtilis]	75	56	702

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
33	12	6878	7183	[gi A00205 FECL	ferredoxin (4Fe-4S) - Clostridium thermacetium	75	56	306
36	1	181	2	[gi 2088739	(AF003141) strong similarity to the FABP/P2/CRBP/CRABP family of transporters (Caenorhabditis elegans)	75	43	180
38	22	14510	15379	[gi 1574058	hypothetical (Haemophilus influenzae)	75	56	870
48	33	23398	24066	[gi 1930092	outer membrane protein (Campylobacter jejuni)	75	56	669
51	1	2	319	[gi 43985	nifs-like gene (Lactobacillus delbrueckii)	75	55	318
51	110	8318	11683	[gi 5337192	CG Site No. 620; alternate gene names hs, hsp, hsr, rms apparent frameshift in GenBank Accession Number X06545 (Escherichia coli)	75	50	3366
54	118	19566	20759	[gi 666069	orf2 gene product (Lactobacillus leichmannii)	75	58	1194
57	9	8448	7822	[gi 290561	ori88 (Escherichia coli)	75	50	627
65	14	6072	6356	[gi 606241	305 ribosomal subunit protein S14 (Escherichia coli)	75	64	285
70	4	3071	2472	[gi 1256617	adenine phosphoribosyltransferase (Bacillus subtilis)	75	57	600
71	24	30399	29404	[gi 1574390	CG-dicarboxylate transport protein (Haemophilus influenzae)	75	57	996
73	2	910	455	[gnl PID e249656	YneT (Bacillus subtilis)	75	57	456
79	1	1810	491	[gi 1146219	28.2% of identity to the Escherichia coli GTP-binding protein Era; putative (Bacillus subtilis)	75	59	1320
82	6	6360	6536	[gi 1655715	Bztd (Rhodobacter capsulatus)	75	55	177
83	6	1938	2975	[gnl PID e123529	putative Plax protein (Bacillus subtilis)	75	56	1018
93	11	7368	5317	[gi 39989	methionyl-tRNA synthetase (Bacillus stearothermophilus)	75	58	2052
93	13	9409	8699	[gi 1591493	glutamine transport ATP-binding protein Q (Methanococcus jannaschii)	75	54	711
95	1	1795	47	[gnl PID e223510	Ylov protein (Bacillus subtilis)	75	57	1749
103	2	362	1186	[gnl PID e266928	unknown (Mycobacterium tuberculosis)	75	64	825
104	1	691	915	[gi 460026	repressor protein (Streptococcus pneumoniae)	75	54	225
113	5	2951	3883	[gnl PID d101119	ABC transporter subunit (Synecocystis sp.)	75	55	933
121	1	320	1390	[gi 2145131	repressor of class I heat shock gene expression HrcA (Streptococcus mutans)	75	58	1071
127	6	2614	3000	[gi 1500451	M. jannaschii predicted coding region MJ1558 (Methanococcus jannaschii)	75	44	387
137	18	10082	10687	[gi 193116	P-glycoprotein 5 (Entamoeba histolytica)	75	52	606
149	11	8499	9338	[gnl PID d100582	unknown (Bacillus subtilis)	75	55	840

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
151	6	9100	7673	gi 40467	hads polypeptide, part of CfrA family [Citrobacter freundii]	75	57	1428
158	1	986	3	gnl PID e253891	UDP-glucose 4-epimerase [Bacillus subtilis]	75	63	984
172	8	5653	6174	gi 142978	glycerol dehydrogenase [Bacillus stearothermophilus]	75	56	1122
172	9	7139	9730	gnl PID e288456	unknown [Mycobacterium tuberculosis]	75	58	2592
173	1	261	79	gnl PID e236469	C10C5.6 [Caenorhabditis elegans]	75	50	183
185	1	3066	2014	gi 1574806	spermidine/putrescine transport ATP-binding protein (potA) [Haemophilus influenzae]	75	56	1051
191	6	5235	4213	gi 149518	phosphoribosyl anthranilate transferase [Lactococcus lactis]	75	61	1023
226	2	1774	1181	gi 2314588	[AE000642] conserved hypothetical protein [Helicobacter pylori]	75	65	594
231	1	1	153	gi 40173	homolog of E. coli ribosomal protein L21 [Bacillus subtilis]	75	57	153
234	1	2	418	gi 2292259	[AF008220] Ycql [Bacillus subtilis]	75	59	417
279	1	552	151	gi 1119198	unknown protein [Bacillus subtilis]	75	50	402
291	7	3558	3827	gi 40011	ORF17 (AA 1-161) [Bacillus subtilis]	75	48	270
375	2	137	628	gi 410137	ORFX1 [Bacillus subtilis]	75	58	492
6	20	16721	17560	gi 2293323	[AF008220] Ycdl [Bacillus subtilis]	74	53	840
7	6	4682	6052	gi 1356211	PEF12-like protein [Bacillus subtilis]	74	60	1371
18	4	3141	2427	gnl PID d101319	Yvg1 [Bacillus subtilis]	74	54	915
21	6	5885	4800	gi 1072381	glutaryl-aminopeptidase [Lactococcus lactis]	74	59	1086
24	2	739	548	gi 2314762	[AE000655] ABC transporter, permease protein [Ysaeg] [Helicobacter pylori]	74	46	192
25	1	2	167	gnl PID d100932	H2O-forming NADH Oxidase [Streptococcus mutans]	74	63	366
38	18	11432	12964	gi 537034	ORF_0488 [Escherichia coli]	74	57	1533
48	10	8924	6669	gi 1513069	P-type adenosine triphosphatase [Listeria monocytogenes]	74	53	2256
55	11	11964	11401	gnl PID e283110	femD [Staphylococcus aureus]	74	64	564
61	2	1782	427	gi 2293216	[AF008220] putative UDP-N-acetylmuramate-alanine ligase [Bacillus subtilis]	74	55	1356
76	10	9414	8065	gnl PID d101325	YqjB [Bacillus subtilis]	74	54	1350
83	2	666	926	pic CJ1496 C314	hisc homolog - Bacillus subtilis	74	55	261
86	9	8985	8080	gi 681585	prephenate dehydratase [Lactococcus lactis]	74	55	906

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
102	5	5005	5652	gi1143394	OMP-PAPP transferase (Bacillus subtilis)	74	57	648
103	5	4364	1267	gnl PID e32324	Y10M protein (Bacillus subtilis)	74	62	1098
108	7	6864	7592	gnl PID e257631	methyltransferase (Lactococcus lactis)	74	56	729
131	2	478	146	gnl PID d101320	YqgZ (Bacillus subtilis)	74	45	333
133	2	1380	919	gnl PID e313025	hypothetical protein (Bacillus subtilis)	74	60	462
137	9	6167	6787	gnl PID d100479	Na <sup>+</sup> -ATPase subunit D (Enterococcus hirae)	74	53	621
149	4	3008	3883	gnl PID d100581	high level kasamycin resistance (Bacillus subtilis)	74	55	876
157	2	243	824	gi11573773	methylated-DNA-protein-cysteine methyltransferase (dall) (Haemophilus influenzae)	74	68	582
164	6	3515	4249	gi1410131	ORFX7 (Bacillus subtilis)	74	68	735
167	7	5446	5201	gi1413927	lpa-3r gene product (Bacillus subtilis)	74	55	246
171	1	1	1818	gnl PID d102251	beta-galactosidase (Bacillus circulans)	74	62	1818
172	4	1064	2392	gi1466474	cellulose phosphorylase enzyme 11'' (Bacillus stearothermophilus)	74	50	1329
185	1	326	3	gi11573646	Mg(2+) transport ATPase protein C (myc) (SP:P22037) (Haemophilus influenzae)	74	68	324
188	2	1089	2018	gi11573008	ATP dependent translocator homolog (asbA) (Haemophilus influenzae)	74	64	930
189	11	6491	7176	gi11661199	sakacin A production response regulator (Streptococcus mutans)	74	60	684
210	2	520	1287	gi12293207	(AF008220) Y1mQ (Bacillus subtilis)	74	60	768
261	1	836	192	gi1666983	putative ATP binding subunit (Bacillus subtilis)	74	55	645
263	3	1619	3655	gi1663232	Similarity with S. cerevisiae hypothetical 137.7 kD protein in subtelomeric Y' repeat region (Saccharomyces cerevisiae)	74	42	2037
265	2	844	1227	gi149272	Asparaginase (Bacillus licheniformis)	74	64	384
368	1	1	942	gi1603998	unknown (Saccharomyces cerevisiae)	74	39	942
7	16	13357	11921	gnl PID d101324	YqjX (Bacillus subtilis)	73	57	1437
17	10	5706	5449	gnl PID e305362	unnamed protein product (Streptococcus thermophilus)	73	47	258
31	2	522	244	gnl PID d100576	single strand DNA binding protein (Bacillus subtilis)	73	55	279
32	6	5667	6194	gnl PID d101315	YqfG (Bacillus subtilis)	73	58	528
34	15	10281	9790	gnl PID d102151	(AB001684) ORF42C (Chlorella vulgaris)	73	46	492



TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
40	12	9876	9226	[gi 1173517	riboflavin synthase alpha subunit (Actinobacillus pleuropneumoniae)	73	55	651
55	2	3592	839	[gnl PID d101087	cation-transporting ATPase PacL (Synecocystis sp.)	73	60	2754
55	18	17494	16586	[gnl PID e265580	unknown (Mycobacterium tuberculosis)	73	52	909
65	16	7213	7767	[gi 143419	ribosomal protein L6 (Bacillus stearothermophilus)	73	60	555
66	3	3300	3659	[gnl PID e269883	LacF (Lactobacillus casei)	73	52	360
70	10	5557	5733	[gi 857631	envelope protein (Human immunodeficiency virus type 1)	73	60	177
71	4	6133	8262	[gnl PID e322063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	73	45	2130
72	1	3	851	[gi 2293177	(AF008220) transporter (Bacillus subtilis)	73	50	849
76	7	7019	6195	[gnl PID d101325	YqjF (Bacillus subtilis)	73	66	825
76	12	10009	9533	[gi 1573086	uridine kinase (uridine monophosphokinase) (udk) (Haemophilus influenzae)	73	54	477
80	7	8113	9372	[gi 377823	aminopeptidase (Bacillus subtilis)	73	60	1260
97	5	3389	1668	[gnl PID d101954	dihydroxyacid dehydratase (Synecocystis sp.)	73	54	1722
98	9	6912	7619	[gnl PID e314991	PisE (Mycobacterium tuberculosis)	73	54	708
108	11	10928	10440	[gi 388109	regulatory protein (Enterococcus faecalis)	73	54	489
128	6	3632	4222	[gi 1685111	orf1091 (Streptococcus thermophilus)	73	63	591
138	2	1575	394	[gi 147326	transport protein (Escherichia coli)	73	60	1182
140	13	12538	11903	[pir E53402 E534	serine O-acetyltransferase (EC 2.3.1.30) - Bacillus stearothermophilus	73	55	636
162	5	5701	4991	[gnl PID e323511	putative YhaO protein (Bacillus subtilis)	73	50	711
164	4	2223	2790	[gi 1592076	hypothetical protein (SP.P25768) (Methanococcus jannaschii)	73	52	468
164	8	4815	5546	[gi 410137	ORFX1 (Bacillus subtilis)	73	56	732
170	5	4394	5302	[gnl PID d100959	homologue of unidentified protein of E. coli (Bacillus subtilis)	73	46	909
178	7	3893	4855	[gi 46242	modulation protein B, 5' end (Rhizobium loti)	73	56	963
204	6	5096	4278	[gnl PID e214719	PicR protein (Bacillus thuringiensis)	73	41	819
213	2	812	2037	[gi 1565296	ribosomal protein S1 homology; sequence specific DNA binding protein (Leuconostoc lactis)	73	55	1206
231	2	84	287	[gi 40173	homolog of E. coli ribosomal protein L21 (Bacillus subtilis)	73	61	204
237	1	2	505	[gi 1773151	adenine phosphoribosyltransferase (Escherichia coli)	73	51	504

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
95	2	2004	1717	gnl PID e323527	putative Asp23 protein [Bacillus subtilis]	72	40	288
109	1	1452	118	gii 143331	alkaline phosphatase regulatory protein [Bacillus subtilis]	72	52	1335
126	1	3	2192	gnl PID d101831	glutamine-binding periplasmic protein [Synechocystis sp.]	72	46	2190
130	3	1735	2478	gii 2415396	[AF015775] carboxypeptidase [Bacillus subtilis]	72	53	744
137	6	2585	2929	gii 472922	V-type Na-ATPase [Enterococcus hirae]	72	46	345
140	10	9601	9203	gii 49224	urf 4 [Synechococcus sp.]	72	48	399
146	5	1906	1247	gnl PID e324945	hypothetical protein [Bacillus subtilis]	72	45	660
147	2	2084	1081	gnl PID e325016	hypothetical protein [Bacillus subtilis]	72	56	1002
147	5	6156	5146	gii 472327	TTP-dependent acetoin dehydrogenase beta-subunit [Clostridium magnum]	72	56	1011
148	8	5381	6433	gii 974332	NAD(P)H-dependent dihydroxyacetone-phosphate reductase [Bacillus subtilis]	72	54	1053
148	14	10256	9675	gnl PID d101319	YggN [Bacillus subtilis]	72	50	582
159	8	4005	4949	gii 1768770	[AE000330] 0463; 24 pct identical (44 gaps) to 338 residues from penicillin-binding protein 4*, PBP4_BACSU SM: P12959 (451 aa) [Escherichia coli]	72	43	945
172	10	9907	10620	gii 761387	unknown [Saccharomyces cerevisiae]	72	55	714
220	3	2862	3602	gii 1574175	hypothetical [Haemophilus influenzae]	72	50	741
267	1	3	449	gii 290511	[470] [Escherichia coli]	72	48	447
281	2	899	540	gnl PID d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis [Bacillus subtilis]	72	45	360
290	1	1018	14	gii 474195	This ORF is homologous to a 40.0 kd hypothetical protein in the htrB 3' region from E. coli. Accession Number X61000 [Mycoplasma-like organism]	72	54	1005
300	1	63	587	gii 746399	transcription elongation factor [Escherichia coli]	72	50	525
316	1	1326	4	gii 158127	protein kinase C [Drosophila melanogaster]	72	40	1323
342	1	227	3	gnl PID d101164	unknown [Bacillus subtilis]	72	54	225
354	1	1	1005	gnl PID d102048	C. thermophilum beta-glucosidase; P26208 (985) [Bacillus subtilis]	72	52	1005
6	10	8134	10467	gnl PID e264229	unknown [Mycobacterium tuberculosis]	71	57	2334
7	20	16231	15464	gii 18046	3-oxoacyl-(acyl-carrier protein) reductase [Cuphea lanceolata]	71	52	768
15	1	1297	2	gnl PID d100571	replicative DNA helicase [Bacillus subtilis]	71	51	1296
15	4	4435	3869	gii 499384	orf189 [Bacillus subtilis]	71	47	567

TABLE 2

S. pneumoniae: Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
18	6	5120	4218	gnl PID d101310	yggG (Bacillus subtilis)	71	51	903
29	1	1	540	gnl 1773142	similar to the 20.2kd protein in TETB-EXOA region of B. subtilis (Escherichia coli)	71	56	540
38	120	11327	113830	gnl 537036	ORF_0158 (Escherichia coli)	71	48	504
51	112	115015	112676	gnl 149528	dipeptidyl peptidase IV (Lactococcus lactis)	71	55	2340
5	123	21040	120585	gnl 2343285	(IAF015433) surface located protein (Lactobacillus rhamnosus)	71	58	456
60	2	705	265	gnl PID d101320	yggZ (Bacillus subtilis)	71	44	441
71	18	24679	26226	gnl 580920	rodD (gtaA) polypeptide (AA 1-673) (Bacillus subtilis)	71	44	1548
71	125	30587	30360	gnl 606028	ORF_0414; Geneplot suggests frameshift near start but none found (Escherichia coli)	71	50	228
72	6	5239	6729	gnl 580835	lysine decarboxylase (Bacillus subtilis)	71	48	1491
72	14	11991	12878	gnl 624085	similar to rat beta-alanine synthetase encoded by GenBank Accession Number S27881; contains ATP/GTP binding motif (Paramacium bursaria Chlorella virus 1)	71	54	888
73	11	7269	7033	gnl 1906594	PM1 (Rattus norvegicus)	71	42	237
74	6	10385	8517	gnl 1573733	prolyl-L-lysine synthetase (proS) (Haemophilus influenzae)	71	52	1869
81	9	5772	6578	gnl 147404	manganese permease subunit II-M-pan (Escherichia coli)	71	45	807
86	5	4602	3604	gnl PID c122063	ss-1,4-galactosyltransferase (Streptococcus pneumoniae)	71	53	999
105	4	3619	4707	gnl 2323341	(IAF014460) PepQ (Streptococcus mutans)	71	58	1089
106	13	13557	12955	gnl 1519287	Lema (Listeria monocytogenes)	71	48	603
114	2	1029	1979	gnl 330303	mosa (Rhizobium meliloti)	71	55	951
122	2	564	1205	gnl 1649037	glutamine transport ATP-binding protein GLNQ (Salmonella typhimurium)	71	50	642
132	5	9018	7063	gnl PID d102049	H. influenzae hypothetical ABC transporter; P44808 (974) (Bacillus subtilis)	71	51	1956
140	1	1141	227	gnl 1633788	(IAE000015) Mycoplasma pneumoniae, fructose-bisphosphate aldolase; similar to Swiss-Prot Accession Number P13243, from B. subtilis (Mycoplasma pneumoniae)	71	49	915
140	5	5635	4973	gnl PID d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus (Bacillus subtilis)	71	48	663
141	7	7369	7845	gnl PID d102005	(IAB001488) FUNCTION UNKNOWN, SIMILAR PRODUCT IN E. COLI AND MYCOPLASMA PNEUMONIAE (Bacillus subtilis)	71	51	477

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
193	1	1	165	gi 46912	ribosomal protein L13 (Staphylococcus carnosus)	71	59	165
194	3	2205	1594	gi 535351	CodY (Bacillus subtilis)	71	52	612
199	3	1510	1319	gi 2182574	(AE000090) Y4pE (Rhizobium sp. MCR234)	71	45	192
208	2	2616	3752	gi 1787378	(AE000311) hypothetical protein in purB 5' region (Escherichia coli)	71	57	1137
209	2	2022	1141	gi 41432	fepC gene product (Escherichia coli)	71	46	882
210	5	1911	3071	gi 49316	ORF2 gene product (Bacillus subtilis)	71	45	1161
210	6	3059	3386	gi 580900	(JMF) gene product (Bacillus subtilis)	71	48	318
212	2	3561	1381	gi 557567	ribonucleotide reductase R1 subunit (Mycobacterium tuberculosis)	71	53	2181
231	3	2003	2920	gnl pid d101320	YqgR (Bacillus subtilis)	71	50	918
244	1	13	1053	gnl pid d100964	homologue of aspartokinase 2 alpha and beta subunits LysC of B. subtilis (Bacillus subtilis)	71	55	1041
251	2	1008	1874	gi 755601	unknown (Bacillus subtilis)	71	46	867
282	2	906	712	gi 1353874	unknown (Rhodobacter capsulatus)	71	46	195
312	4	2117	1565	gnl pid d102245	(AB005554) Y2bF (Bacillus subtilis)	71	34	573
338	1	3	683	gi 1591045	hypothetical protein (SP:P31466) (Methanococcus jannaschii)	71	48	681
346	1	3	164	gi 1591234	hypothetical protein (SP:P42297) (Methanococcus jannaschii)	71	36	162
374	1	619	2	gi 397526	clumping factor (Staphylococcus aureus)	71	23	618
377	1	688	2	gi 397526	clumping factor (Staphylococcus aureus)	71	23	687
3	8	7419	6958	gnl pid e269486	unknown (Bacillus subtilis)	70	42	462
3	10	8395	9075	gnl pid e25543	putative iron dependant repressor (Staphylococcus epidermidis)	70	46	681
7	14	11024	10254	gnl pid d100290	undefined open reading frame (Bacillus stearothermophilus)	70	55	771
7	18	14213	13719	gnl pid d101090	biotin carboxyl carrier protein of acetyl-CoA carboxylase (Synechocystis sp.)	70	56	495
9	2	1057	287	gnl pid d100581	unknown (Bacillus subtilis)	70	52	771
12	4	2610	1789	gnl pid d101195	lycJ (Bacillus subtilis)	70	52	822
21	2	2586	1846	gi 2293447	(AF008930) ATPase (Bacillus subtilis)	70	54	741
22	13	10955	11512	gi 1165295	Ydr540cp (Saccharomyces cerevisiae)	70	50	558
30	6	4315	1980	gi 39478	ATP binding protein of transport ATPases (Bacillus firmus)	70	51	336

TABLE 2  
S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
31	1	370	113	[gi 662792	[single-stranded DNA binding protein (unidentified eubacterium)	70	36	258
33	15	10639	9521	[gi 1161219	[homologous to D-amino acid dehydrogenase enzyme (Pseudomonas aeruginosa)	70	50	1119
38	6	3812	4312	[gi 2058547	[CusM (Streptococcus gordonii)	70	48	501
38	25	17986	18477	[gi 537033	[ORF f356 (Escherichia coli)	70	58	492
40	13	11054	9846	[gi 1173516	[riboflavin-specific deaminase (Actinobacillus pleuropneumoniae)	70	52	1209
42	2	722	1954	[gi 1146183	[putative (Bacillus subtilis)	70	51	1233
43	3	2373	1612	[gi 1591493	[glutamine transport ATP-binding protein Q (Methanococcus jannaschii)	70	48	762
45	8	9197	8049	[gnl pid d102036	[subunit of ADP-glucose pyrophosphorylase (Bacillus stearothermophilus)	70	54	1149
59	2	567	956	[gnl pid d100302	[neopullulanase (Bacillus sp.)	70	42	390
60	3	1874	795	[gnl pid e276466	[aminopeptidase P (Lactococcus lactis)	70	48	1080
61	4	5553	2437	[gnl pid e275074	[SNF (Bacillus cereus)	70	51	3117
61	7	7914	6802	[gi 1573037	[cystathionine gamma-synthase (malb) (Haemophilus influenzae)	70	52	1113
63	7	5372	7222	[gnl pid d100974	[unknown (Bacillus subtilis)	70	54	1851
68	7	7126	6962	[gi 1263014	[ema18.1 gene product (Streptococcus pyogenes)	70	37	165
72	12	10081	10911	[gi 2313093	[AAC00524) carboxymethyltransferase (nspC) (Helicobacter pylori)	70	56	831
75	10	7888	8124	[gi 1877423	[galactose-1-P-uridylyl transferase (Streptococcus mutans)	70	59	237
79	3	3424	2525	[gi 139881	[ORF 311 (AA 1-311) (Bacillus subtilis)	70	47	900
87	10	9369	7324	[gnl pid e123506	[putative Pkn2 protein (Bacillus subtilis)	70	52	2046
96	14	10640	11788	[gi 1573209	[tRNA guanine transglycosylase (tgt) (Haemophilus influenzae)	70	52	1149
113	2	574	1086	[gi 433630	[A160 (Saccharomyces cerevisiae)	70	59	513
123	5	2901	3461	[gnl pid d100585	[unknown (Bacillus subtilis)	70	45	561
125	5	4593	4282	[gnl pid e276474	[capacitative calcium entry channel 1 (Bos taurus)	70	35	312
129	5	4500	3454	[gnl pid d101314	[yqet (Bacillus subtilis)	70	47	1047
131	3	2608	1394	[gi 2293312	[AAC00220) YCFP (Bacillus subtilis)	70	50	1215
135	1	420	662	[gnl pid e265530	[yofE (Streptococcus pneumoniae)	70	47	243
137	3	438	932	[gi 472919	[v-type H <sup>+</sup> -ATPase (Enterococcus hirae)	70	57	495
138	1	440	3	[gi 147336	[transmembrane protein (Escherichia coli)	70	42	438

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
140	16	118796	16164	gi1976441	MS-methyltetrahydrofolate homocysteine methyltransferase [Saccharomyces cerevisiae]	70	53	2433
167	10	8263	6695	gi1149535	D-alanine activating enzyme [Lactobacillus casei]	70	52	1569
204	4	3226	2747	gnl PID d102049	E. coli hypothetical protein; P11805 (267) [Bacillus subtilis]	70	51	480
207	3	2627	2869	gnl PID e309213	racGAP [Dictyostellium discoideum]	70	45	243
282	3	1136	882	gi11353874	unknown [Rhodobacter capsulatus]	70	50	255
6	21	17554	18453	gnl PID e233879	hypothetical protein [Bacillus subtilis]	69	44	900
6	22	118482	19471	gi1580883	jipa-88d gene product [Bacillus subtilis]	69	53	990
22	6	4682	5824	gi12209379	[AF006720] ProJ [Bacillus subtilis]	69	48	1143
22	9	7992	8651	gnl PID d100580	unknown [Bacillus subtilis]	69	51	660
22	12	9871	10767	gnl PID d100581	unknown [Bacillus subtilis]	69	51	897
27	7	5857	5348	gnl PID d102012	[AB001488] FUNCTION UNKNOWN [Bacillus subtilis]	69	28	510
36	10	7294	10116	gi1437916	isoleucyl-tRNA synthetase [Staphylococcus aureus]	69	53	2823
38	1	2	1090	gi1141900	alcohol dehydrogenase (EC 1.1.1.1) [Alcaligenes eutrophus]	69	48	1089
40	14	11333	11944	gi11573280	Holliday junction DNA helicase (ruvA) [Haemophilus influenzae]	69	44	612
40	15	11142	12517	gi11573653	DNA-3-methyladenine glycosidase 1 (tagI) [Haemophilus influenzae]	69	50	576
45	6	6947	5490	gi1580887	starch (bacterial glycogen) synthase [Bacillus subtilis]	69	47	1458
48	14	124932	24153	gnl PID e233870	hypothetical protein [Bacillus subtilis]	69	36	780
49	6	6181	6521	gi1396297	similar to phosphotransferase system enzyme II [Escherichia coli]	69	50	339
49	8	7586	8138	gi1396420	similar to Alcaligenes eutrophus pncI D-ribulose-5-phosphate 3 epimerase [Escherichia coli]	69	49	753
55	6	8262	7033	gi11146238	[poly(A) polymerase [Bacillus subtilis]	69	50	1230
59	3	954	2333	gnl PID e113038	hypothetical protein [Bacillus subtilis]	69	54	1380
62	3	1170	1418	gnl PID d101915	hypothetical protein [Synecocystis sp.]	69	49	249
63	8	7298	7762	gi1293017	[ORF] (put. 1): putative [Lactococcus lactis]	69	42	465
66	4	3657	5081	gi1153755	phospho-beta-D-galactosidase (EC 3.2.1.85) [Lactococcus lactis cremoris]	69	49	1425
66	5	5126	6829	gi1433809	penzyme II [Streptococcus mutans]	69	46	1704
71	6	10017	10664	gnl PID e322061	[ss-1,4-galactosyltransferase [Streptococcus pneumoniae]	69	39	648

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
71	21	27730	27966	gnt PIU d100649	DE-cadherin (Drosophila melanogaster)	69	30	237
77	1	1	237	g1 287870	groES gene product (Lactococcus lactis)	69	44	237
81	5	3622	4101	g1 1513605	fucose operon protein (fucU) (Haemophilus influenzae)	69	52	480
83	1	40	714	pir CJ1496 C334	hisc homolog - Bacillus subtilis	69	46	675
83	16	15742	16335	g1 143372	phosphoribosyl glycineamide formyltransferase (pur-N) (Bacillus subtilis)	69	46	594
85	2	1212	916	g1 194097	IFN-response element binding factor 1 (Mus musculus)	69	48	297
91	5	3678	4274	g1 1574712	anaerobic ribonucleoside-triphosphate reductase activating protein (nrdG) (Haemophilus influenzae)	69	44	597
98	5	3247	4032	gnt PIU d100262	livF protein (Salmonella typhimurium)	69	51	786
108	5	4085	5056	gnt PIU e257629	transcription factor (Lactococcus lactis)	69	49	972
126	3	3078	4568	gnt PIU d101329	YqjJ (Bacillus subtilis)	69	49	1491
131	6	4121	2889	gnt PIU d101314	YqeR (Bacillus subtilis)	69	47	1233
136	2	1505	2299	gnt PIU d100581	unknown (Bacillus subtilis)	69	47	795
149	5	3852	4763	gnt PIU e323525	YloQ protein (Bacillus subtilis)	69	50	932
149	12	9336	10655	g1 151571	homology with E.coli and P.aeruginosa lysA gene; product of unknown function; putative (Pseudomonas syringae)	69	52	1320
153	4	3191	3829	g1 1710373	BrnQ (Bacillus subtilis)	69	44	639
169	3	849	2124	gnt PIU d100582	temperature sensitive cell division (Bacillus subtilis)	69	49	1476
180	1	566	3	g1 488339	alpha-amyase (unidentified cloning vector)	69	50	564
212	1	1196	231	g1 1395209	ribonucleotide reductase R2-2 small subunit (Mycobacterium tuberculosis)	69	53	966
226	1	2	661	pir JQ2285 JQ22	nodulin-26 - soybean	69	41	660
233	5	3249	4766	g1 472918	v-type Na-ATPase (Enterococcus hirae)	69	56	1518
235	3	660	1766	g1 148945	methylase (Haemophilus influenzae)	69	43	1107
243	2	865	2161	gnt PIU d100225	OMF5 (Barley yellow dwarf virus)	69	69	1497
251	3	2899	1967	g1 2289231	macrolide-efflux protein (Streptococcus agalactiae)	69	51	933
310	1	1	282	gnt PIU e122442	peptide deformylase (Clostridium beijerinckii)	69	55	282
369	1	668	2	g1 397526	clumping factor (Staphylococcus aureus)	69	22	867
370	1	749	3	g1 397526	clumping factor (Staphylococcus aureus)	69	21	747

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
379	1	44	280	[gnl PID d100649]	DE-cadherin (Drosophila melanogaster)	69	30	237
388	1	260	72	[gi 11787524]	[AE000225] hypothetical 32.7 kD protein in trpL-burA intergenic region [Escherichia coli]	69	44	189
1	2	2006	3040	[gnl PID d101809]	ABC transporter [Synecocystis sp.]	68	43	1035
12	5	3958	2600	[gi 2182992]	histidine kinase (Lactococcus lactis cremoris)	68	45	1359
15	2	1790	1311	[pir S16974 R585]	ribosomal protein L9 - Bacillus stearothermophilus	68	56	480
16	6	7353	5701	[gi 1787041]	[AE000184] 0530: This 530 aa ort is 33 pct identical (14 gaps) to 525 residues of an approx. 640 aa protein YHE5_MAEIN SW: P44808 [Escherichia coli]	68	45	1653
17	12	6479	6805	[gi 553165]	acetylcholinesterase (Homo sapiens)	68	68	327
20	13	14128	14505	[gi 142700]	p competence protein (ttg start codon) (put. 1: putative [Bacillus subtilis])	68	40	378
22	32	124612	25197	[gi 289262]	[comE ORF3] [Bacillus subtilis]	68	36	786
30	7	4548	4288	[gi 311388]	ORF1 [Azorhizobium caulinodans]	68	46	261
36	5	3911	4585	[gi 1573041]	hypothetical [Haemophilus influenzae]	68	54	675
46	6	5219	6040	[gi 1790131]	[AE000446] hypothetical 29.7 kD protein in tbaA-yrbB intergenic region [Escherichia coli]	68	47	822
54	10	6235	7086	[gi 882579]	CC Site No. 29739 [Escherichia coli]	68	55	852
55	5	7069	5165	[gnl PID d101914]	ABC transporter [Synecocystis sp.]	68	45	1905
71	3	6134	5613	[gi 1573353]	outer membrane integrity protein (tolA) [Haemophilus influenzae]	68	50	522
71	10	15342	16613	[gi 580866]	ipa-12d gene product [Bacillus subtilis]	68	31	1272
71	12	17560	18792	[gi 44073]	[SecY protein [Lactococcus lactis]	68	35	1233
71	17	22295	24703	[gi 1762349]	involved in protein export [Bacillus subtilis]	68	50	2409
73	16	10208	9729	[gi 1353537]	[dUTPase [Bacteriophage phi]	68	51	480
86	18	17198	16011	[gi 411943]	ipa-19d gene product [Bacillus subtilis]	68	53	1188
87	17	17491	15866	[gi 150209]	[ORF 1] [Mycoplama mycoides]	68	43	1626
89	6	5139	4154	[gi 1498824]	M. jannaschii predicted coding region NJ0062 [Methanococcus jannaschii]	68	40	786
89	11	8021	8242	[gi 150974]	[4-oxalocrotonate tautomerase [Pseudomonas putida]	68	43	222
97	8	6755	5194	[gi 2367358]	[AE000491] hypothetical 52.9 kD protein in atdB-rpsF intergenic region [Escherichia coli]	68	41	1362



TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
98	3	1418	2108	[gnl PID d100261]	LtVA protein [Salmonella typhimurium]	68	40	891
99	11	16414	17280	[gi 455363]	regulatory protein [Streptococcus mutans]	68	50	867
115	3	5034	3693	[gi 466474]	cellulose phosphotransferase enzyme II' [Bacillus stearothermophilus]	68	44	1362
124	7	3394	3221	[gnl PID d100702]	cutA protein [Schizosaccharomyces pombe]	68	56	174
125	2	2923	1922	[gi 450366]	transmembrane protein [Bacillus subtilis]	68	50	1002
132	2	4858	2888	[gnl PID d101732]	DNA ligase [Synecocystis sp.]	68	52	1971
140	7	7765	7580	[gi 1209711]	unknown [Saccharomyces cerevisiae]	68	47	186
150	1	519	3	[gi 402490]	ADP-ribosylarginine hydrolase [Mus musculus]	68	59	537
164	1	58	867	[gnl PID e255114]	glutamate racemase [Bacillus subtilis]	68	49	810
170	2	819	1835	[gnl PID e255117]	hypothetical protein [Bacillus subtilis]	68	50	1017
179	7	3946	4104	[pir B54545 B545]	hypothetical protein - Lactococcus lactis subsp. lactis plasmid pSL2	68	40	159
170	4	4247	4396	[gi 304146]	spore coat protein [Bacillus subtilis]	68	52	150
171	8	6002	7054	[gi 38722]	precursor (aa 20 to 381) [Acinetobacter calcoaceticus]	68	54	1053
198	3	2473	1871	[gnl PID e13075]	hypothetical protein [Bacillus subtilis]	68	46	603
211	2	969	1802	[gi 1439528]	E1C-man [Lactobacillus curvatus]	68	45	834
214	8	4926	4231	[gnl PID d102049]	M. influenzae hypothetical protein; P43990 (182) [Bacillus subtilis]	68	50	696
217	6	4955	5170	[gnl PID e326966]	similar to B. vulgaris CMS-associated mitochondrial (reverse transcriptase) [Arabidopsis thaliana]	68	36	216
218	7	3930	4745	[gi 2293198]	[AF008220] YtgP [Bacillus subtilis]	68	38	816
220	6	4628	4318	[gnl PID e125791]	[AJ000005] orf1 [Bacillus megaterium]	68	51	291
236	3	746	108	[gi 410137]	ORFX13 [Bacillus subtilis]	68	46	639
237	2	675	1451	[gi 396348]	homoserine transsuccinylase [Escherichia coli]	68	49	777
250	4	771	1229	[gi 310859]	ORF2 [Synecococcus sp.]	68	50	459
254	1	517	155	[gi 1787105]	[AE000189] o648 was o669; This 669 aa orf is 40 pct identical (1 gap) to 217 residues of an approx. 232 aa protein Y8A_HAEIN SM: P45247 [Escherichia coli]	68	44	363
337	1	1	774	[gnl PID e261990]	putative orf [Bacillus subtilis]	68	47	774
345	1	3	653	[gi 149513]	[thymidylate synthase (EC 2.1.1.45) [Lactococcus lactis]	68	61	651

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
386	2	417	4	[gi1573353]	outer membrane integrity protein (tolA) [Haemophilus influenzae]	68	51	414
2	4	5222	4697	[gi1592141]	M. jannaschii predicted coding region MJ1507 [Methanococcus jannaschii]	67	26	1026
3	6	5397	4591	[gi12293175]	[AF008220] signal transduction regulator [Bacillus subtilis]	67	44	807
5	2	2301	574	[gi12313385]	[AE00547] para-aminobenzoate synthetase (pabB) [Helicobacter pylori]	67	48	1728
6	19	16063	16758	[gi1413931]	lipa-7d gene product [Bacillus subtilis]	67	41	696
22	8	7094	7897	[gi11928962]	pyrroline-5-carboxylate reductase [Actinidia deliciosa]	67	51	804
29	10	8335	9072	[gi1468745]	gtcr gene product [Bacillus brevis]	67	41	738
31	3	1379	585	[gi12425123]	[AF019986] PabB [Dictyostelium discoideum]	67	49	795
32	11	8849	10150	[gi142029]	ORF1 gene product [Escherichia coli]	67	47	1302
36	16	114810	15546	[gi1592142]	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	67	43	717
38	9	4958	5392	[gn11P1D]e214803	T2283.3 [Caenorhabditis elegans]	67	47	435
38	21	11375	14512	[gi1537017]	ORF_0216 [Escherichia coli]	67	52	738
45	9	11028	9181	[gi1551710]	branching enzyme (glgB) [EC 2.4.1.18] [Bacillus stearothermophilus]	67	51	1268
48	23	18344	17514	[gi1413949]	lipa-25d gene product [Bacillus subtilis]	67	50	831
50	2	1773	352	[gn11P1D]d101330	[YqJ0] [Bacillus subtilis]	67	55	822
53	1	431	3	[gi1574291]	[flaB] transcription regulation repressor (pilB) [Haemophilus influenzae]	67	40	429
55	13	112740	111946	[gn11P1D]e252990	ORF YDL037c [Saccharomyces cerevisiae]	67	51	795
61	9	9210	8329	[gn11P1D]e264711	ATP-binding cassette transporter A [Staphylococcus aureus]	67	50	882
71	2	5614	6117	[gi1197667]	[vitellogenin (Anolis pulchellus)]	67	36	504
81	7	4489	4983	[gi1142714]	[phosphoenolpyruvate:mannose phosphotransferase element 118] [Lactobacillus curvatus]	67	42	495
83	7	2957	3214	[gi11276746]	Acyl carrier protein [Porphyra purpurea]	67	37	258
86	8	8140	6809	[gi1147744]	PSR [Enterococcus hirae]	67	45	1332
97	3	986	1366	[gn11P1D]d102235	[AB000631] unnamed protein product [Streptococcus mutans]	67	43	381
102	1	601	1413	[gi1682765]	[mcc8 gene product] [Escherichia coli]	67	36	813
106	3	1109	1987	[gi1148921]	[LicD protein] [Haemophilus influenzae]	67	43	879
115	4	5982	5656	[gi1895750]	[putative cellobiose phosphotransferase enzyme 111] [Bacillus subtilis]	67	44	327

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
115	7	8421	8077	gi 466473	cellobiose phosphotransferase enzyme II [Bacillus stearothermophilus]	67	51	345
127	113	8127	7021	gi 147326	transport protein [Escherichia coli]	67	45	1107
136	3	2215	2859	gnl PID100581	unknown [Bacillus subtilis]	67	49	645
140	21	23317	20906	gnl PID101912	phenylalanine-tRNA synthetase [Synchocystis sp.]	67	43	2412
146	6	2894	1893	gi 2182994	histidine kinase [Lactococcus lactis cremoris]	67	44	1002
151	8	11476	11117	gnl PID100085	ORF129 [Bacillus cereus]	67	48	360
160	10	7453	8646	gi 2281317	OrfB, similar to a Streptococcus pneumoniae putative membrane protein encoded by GenBank Accession Number X99400; inactivation of the OrfB gene leads to UV-sensitivity and to decrease of homologous recombination (plasmidic test) [Lactococcus l.	67	46	1194
163	3	3099	4505	gnl PID101317	YqfR [Bacillus subtilis]	67	47	1407
167	8	6704	5454	gi 1161933	DltB [Lactobacillus casei]	67	45	1251
169	4	2322	2879	gnl PID101331	YqkG [Bacillus subtilis]	67	41	558
171	11	7656	8184	gi 153841	pneumococcal surface protein A [Streptococcus pneumoniae]	67	50	729
188	3	1930	3723	gi 1542975	AbcB [Thermoanaerobacterium thermosulfurigenes]	67	46	1794
189	6	3599	3141	gnl PID10325178	hypothetical protein [Bacillus subtilis]	67	52	459
205	3	1663	2211	gi 606073	ORF_0169 [Escherichia coli]	67	47	549
207	4	2896	3456	gi 2276374	DtxR/iron regulated lipoprotein precursor [Corynebacterium diphtheriae]	67	49	561
217	3	4086	3703	gi 895750	putative cellobiose phosphotransferase enzyme III [Bacillus subtilis]	67	42	384
246	2	291	662	gi 1842438	unknown [Bacillus subtilis]	67	43	372
252	1	2	745	gi 2351768	PspA [Streptococcus pneumoniae]	67	41	744
265	3	1134	1811	gi 2313847	(AE000585) L-asparaginase II (aneB) [Helicobacter pylori]	67	42	678
295	1	1	375	gi 2276374	DtxR/iron regulated lipoprotein precursor [Corynebacterium diphtheriae]	67	43	375
3	7	4898	5146	gnl PID10255179	unknown [Mycobacterium tuberculosis]	66	56	249
3	1	189	3	gnl PID10269548	unknown [Bacillus subtilis]	66	48	387
3	20	19267	20805	gi 39956	lIGlc [Bacillus subtilis]	66	50	1539
4	3	2545	2718	gi 1787564	(AE000228) phage shock protein C [Escherichia coli]	66	36	174
5	9	11197	12592	gi 1574291	fimbrial transcription regulation repressor (pilB) [Haemophilus influenzae]	66	46	606

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
9	4	2872	1451	gnl PID e266928	unknown (Mycobacterium tuberculosis)	66	43	1422
12	2	1469	1200	gi 520407	orf2: GTG start codon (Bacillus thuringiensis)	66	42	270
15	12	10979	9897	gi 2314738	(AE000653) translation elongation factor EF-Ts (tsf) (Helicobacter pylori)	66	49	1083
16	2	1312	734	gnl PID d102245	(AB005554) ynfF (Bacillus subtilis)	66	35	579
22	3	1372	1851	gi 1480916	signal peptidase type II (Lactococcus lactis)	66	38	480
22	7	5828	7096	gnl PID e206261	gamma-glutamyl phosphate reductase (Streptococcus thermophilus)	66	51	1269
22	20	16194	17138	gnl PID e281914	YnfL (Bacillus subtilis)	66	50	945
30	2	530	976	gi 2314379	(AE000627) ABC transporter, ATP-binding protein (yhcG) (Helicobacter pylori)	66	40	447
32	1	199	984	gi 312444	ORF2 (Bacillus caldolyticus)	66	49	786
33	13	8352	7234	gi 1387979	44% identity over 302 residues with hypothetical protein from Synchocystis sp. accession D64006.CD; expression induced by environmental stress; some similarity to glycosyl transferases; two potential membrane-spanning helices (Bacillus subtilis)	66	44	1119
34	6	5658	4708	gnl PID e250724	orf2 (Lactobacillus sake)	66	39	951
34	14	9792	9574	gi 1590997	M. jannaschii predicted coding region MJ0772 (Methanococcus jannaschii)	66	48	219
35	16	15163	14501	gi 1773352	CapSM (Staphylococcus aureus)	66	46	663
36	9	6173	6976	gi 1518680	minicell associated protein BlvIVA (Bacillus subtilis)	66	35	804
36	11	110396	10824	bus 155144	insulin activator factor, INSAF (human, Pancreatic insulinoma, Peptide Partial, 744 aa) (Homo sapiens)	66	43	429
48	1	28	1419	gnl PID e325204	hypothetical protein (Bacillus subtilis)	66	50	1392
48	7	3810	4112	gi 2182574	(AE000090) Y4P6 (Rhizobium sp. NGR234)	66	40	303
52	4	3595	2789	gi 388565	major cell-binding factor (Campylobacter jejuni)	66	52	807
54	3	2662	1076	gnl PID d101811	glutamine-binding periplasmic protein (Synchocystis sp.)	66	43	1587
61	10	9740	9183	gnl PID e154144	mdr gene product (Staphylococcus aureus)	66	44	558
72	13	10893	11193	gi 2313129	(AE000526) H. pylori predicted coding region HP0049 (Helicobacter pylori)	66	44	1101
74	9	13267	12476	gi 1573941	hypothetical (Haemophilus influenzae)	66	43	792
75	1	2	868	gi 1574631	nicotinamide mononucleotide transporter (pnuC) (Haemophilus influenzae)	66	48	867
75	7	5103	4275	gi 41312	put. EBG repressor protein (Escherichia coli)	66	40	1029

TABLE 2  
S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
82	7	6813	8123	gnl PIU e255128	trigler factor (Bacillus subtilis)	66	53	1311
83	3	905	1219	gnl C31496 C336	hinc homology - Bacillus subtilis	66	44	315
86	10	9407	8925	gnl 603584	shikimate kinase (Lactococcus lactis)	66	41	483
88	10	7001	6060	gnl 2098719	putative fimbrial-associated protein (Actinomyces naeslundii)	66	52	942
89	1	951	4	gnl 410118	ORFX19 (Bacillus subtilis)	66	41	948
93	7	3661	2711	gnl 1787936	(AE000260) (298; This 298 aa orf is 51 pct identical (5 gaps) to 297 residues of an approx. 308 aa protein YCSU_BACSU SW: R42972 (Escherichia coli)	66	49	951
104	3	1805	3049	gnl 1469784	putative cell division protein ftsW (Enterococcus hirae)	66	48	1245
106	14	13576	14253	gnl 40027	homologous to E.coli gidB (Bacillus subtilis)	66	52	678
107	3	965	1864	gnl 144858	ORF A (Clostridium perfringens)	66	49	900
112	7	5718	6593	gnl 609332	OpvA (Haemophilus influenzae)	66	43	876
115	1	3	302	gnl 727367	MyrIp (Saccharomyces cerevisiae)	66	56	300
122	1	3	566	gnl PIU d101328	YqiY (Bacillus subtilis)	66	36	564
126	8	11759	11046	gnl PIU d101163	ORF3 (Bacillus subtilis)	66	48	734
128	11	8201	8431	gnl 726288	growth associated protein GAP-43 (Xenopus laevis)	66	41	231
131	8	4894	4508	gnl 486661	TMna related protein (Saccharomyces cerevisiae)	66	39	387
140	3	3236	2574	gnl 40056	phoP gene product (Bacillus subtilis)	66	16	663
140	15	16318	15434	gnl 1658189	5,10-methylenetetrahydrofolate reductase (Erwinia carotovora)	66	48	885
146	12	7926	7636	gnl PIU d101140	transposase (Synecocystis sp.)	66	42	291
147	6	7137	6154	gnl 472326	TPP-dependent acetoin dehydrogenase alpha-subunit (Clostridium magnum)	66	48	984
149	6	4435	5430	gnl PIU d101087	pentose-5-phosphate-3-epimerase (Synecocystis sp.)	66	46	996
149	13	10754	11575	gnl 42371	pyruvate formate-lyase activating enzyme (AA 1-246) (Escherichia coli)	66	42	822
186	4	2578	2270	gnl PIU d101199	ORF11 (Enterococcus faecalis)	66	41	309
207	2	2340	2597	gnl PIU e321893	envelope glycoprotein gp160 (human immunodeficiency virus type 1)	66	46	258
210	7	1358	3678	gnl 49318	ORF4 gene product (Bacillus subtilis)	66	46	321
217	8	5143	5355	gnl 49538	thrombin receptor (Cricetulus longicaudatus)	66	38	213
220	4	3875	3642	gnl 466648	alternate name ORF6 of L23635 (Escherichia coli)	66	33	234

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	Chr	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
223	1	1070	138	[gnl pid e247187]	zinc finger protein [Bacteriophage phigle]	66	45	933
224	2	1864	2640	[gi 1176199]	putative ABC transporter subunit [Staphylococcus epidermidis]	66	41	777
243	1	3	872	[dnj AB000617.2]	[AB000617] YcdH [Bacillus subtilis]	66	45	870
268	2	891	568	[gi 1517210]	putative transposase [Streptococcus pyogenes]	66	60	324
322	1	2	643	[gi 1499836]	[2n protease [Methanococcus jannaschii]	66	40	642
5	10	11309	13178	[gi 1574292]	hypothetical [Haemophilus influenzae]	65	34	732
6	11	10465	11190	[gi 142854]	homologous to E. coli radC gene product and to unidentified protein from Staphylococcus aureus [Bacillus subtilis]	65	48	726
7	2	647	405	[pir C64146 C641]	hypothetical protein III0259 - Haemophilus influenzae (strain Rd KW20)	65	42	243
7	1	6246	6821	[gnl ptu d101323]	[yqhu [Bacillus subtilis]	65	50	576
10	2	1873	1397	[gi 1163111]	ORF-1 [Streptococcus pneumoniae]	65	54	477
16	3	1428	2222	[gnl pid e325010]	hypothetical protein [Bacillus subtilis]	65	45	795
21	4	3815	3357	[gnl pid e314910]	hypothetical protein [Staphylococcus sciuri]	65	40	459
22	14	25776	26384	[gi 1121030]	[CpA [Actinobacillus pleuropneumoniae]	65	42	609
43	2	1648	290	[gi 1044826]	[F14E5.1 [Caenorhabditis elegans]	65	38	1359
48	13	10062	10856	[gi 1573390]	hypothetical [Haemophilus influenzae]	65	45	795
48	22	17521	16883	[gi 1573391]	hypothetical [Haemophilus influenzae]	65	37	619
48	25	119027	18533	[gnl pid e264484]	[YCR020C, Ien.215 [Saccharomyces cerevisiae]	65	38	495
49	3	3856	5334	[gi 1480429]	putative transcriptional regulator [Bacillus stearothermophilus]	65	32	1479
50	6	5337	4519	[gi 171961]	[tRNA isopentenyl transferase [Saccharomyces cerevisiae]	65	42	819
52	15	114728	15588	[gi 1499745]	[M. jannaschii predicted coding region MJ0912 [Methanococcus jannaschii]	65	46	861
59	7	3963	4745	[gi 496514]	[orf zeta [Streptococcus pyogenes]	65	42	783
68	3	2500	3483	[gi 887824]	[ORF_0310 [Escherichia coli]	65	46	984
69	3	2171	1077	[gnl pid e111453]	unknown [Bacillus subtilis]	65	42	1095
69	7	6029	5325	[gi 809660]	[deoxyribose-phosphate aldolase [Bacillus subtilis]	65	55	705
71	5	8536	9783	[gi 1573224]	[glycosyl transferase lgcC [GP:U1454.4] [Haemophilus influenzae]	65	42	1248
72	8	7664	8527	[gnl pid e267589]	Unknown, highly similar to several spermidine synthases [Bacillus subtilis]	65	39	864

TABLE 2

S. pneumoniae - putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
76	5	5773	4097	gnl pid d101723	DNA REPAIR PROTEIN RECN (RECOMBINATION PROTEIN N) [Escherichia coli]	65	44	1677
76	9	8099	7875	gi 1574276	[exodeoxyribonuclease, small subunit (xseB) (Haemophilus influenzae)]	65	38	225
84	2	2870	2352	gi 2313188	[AE000532] conserved hypothetical protein (Helicobacter pylori)	65	41	519
86	15	14495	13407	gnl pid d101880	3-dehydroquinase synthase [Synecocystis sp.]	65	44	1089
87	3	3706	2423	gi 151259	HMG-CoA reductase (EC 1.1.1.88) [Pseudomonas mevalonii]	65	51	1284
88	3	2425	2736	gi 1098510	unknown [Lactococcus lactis]	65	30	312
89	2	1627	1007	gnl pid d102008	[AB001488] SIMILAR TO ORF14 OF ENTEROCOCCUS FAECALIS TRANSPOSON TN916. [Bacillus subtilis]	65	41	621
111	6	6635	6186	gnl pid e246063	[NM23]/nucleoside diphosphate kinase [Xenopus laevis]	65	50	450
116	1	3	1016	gnl pid d101125	[queuosine biosynthesis protein QueA] [Synecocystis sp.]	65	44	1014
123	1	69	389	gi 498839	[ORF2] [Clostridium perfringens]	65	36	321
123	7	6522	7190	gi 1575577	[DNA-binding response regulator [Thermotoga maritima]]	65	39	669
125	3	3821	2859	gnl pid e257609	[sugar-binding transport protein [Anaerocellum thermophilum]]	65	47	963
137	12	8015	7818	gi 2182574	[AE000090] Y4pE [Rhizobium sp. NGR234]	65	41	198
147	4	5021	3885	gi 472329	[dihydrolipoamide acetyltransferase [Clostridium magnum]]	65	47	1137
148	2	1053	1931	gnl pid d101119	[yqgm [Bacillus subtilis]]	65	42	879
151	2	3212	4687	gi 304897	[EcoE type I restriction modification enzyme M subunit [Escherichia coli]]	65	50	1476
156	2	730	437	gi 310893	[membrane protein [Theileria parva]]	65	47	294
164	7	4256	4817	gi 410132	[ORFX8 [Bacillus subtilis]]	65	48	582
169	6	3192	3914	gi 1552737	[similar to purine nucleoside phosphorylase (deoD) [Escherichia coli]]	65	41	723
176	4	2951	2220	gnl pid e339500	[oligopeptide binding lipoprotein [Streptococcus pneumoniae]]	65	43	732
195	4	4556	1900	gi 1592142	[ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]]	65	40	657
196	1	160	1572	gnl pid d102004	[AB001488] PROBABLE UDP-N-ACETYLURACONYLALANYL-D-GLUTAMYL-2, 6-DIAMINOLICASE (EC 6.3.2.15) [Bacillus subtilis]	65	51	1413
204	2	2246	1215	gi 143156	[membrane bound protein [Bacillus subtilis]]	65	37	1032
210	4	1544	1891	gi 49315	[ORF1 gene product [Bacillus subtilis]]	65	48	348
242	2	1625	723	gi 1787540	[AE000226] f249; This 249 aaorf is 32 pct identical (8 gaps) to 244 residues of an approx. 272 aa protein AGAR_ECOLI SM: p42502 [Escherichia coli]	65	42	903

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
284	1	1	900	[gi 559861	[c]ym [Plasmid pAD1]	65	36	900
304	1	2	574	[gnl pt0 e290934	[unknown [Mycobacterium tuberculosis]	65	52	573
315	1	2	1483	[gi 790694	[mannuronan C-5-epimerase [Azotobacter vinelandii]	65	57	1482
320	1	3	569	[gnl pt0 d102048	[K. aerogenes, histidine utilization repressor; p12380 (199) DNA binding [Bacillus subtilis]	65	46	567
358	1	1	309	[gnl pt0 e323508	[Y105 protein [Bacillus subtilis]	65	55	309
2	7	7571	6696	[gi 1498753	[nicotinamide-nucleotide pyrophosphorylase [Rhodospirillum rubrum]	64	47	876
6	6	5924	6802	[gnl pt0 d101111	[methionine aminopeptidase [Synechocystis sp.]	64	52	879
8	4	3417	3686	[gi 1045935	[DNA helicase II [Mycoplasma genitalium]	64	58	270
11	4	3249	2689	[gnl pt0 e265529	[orfB [Streptococcus pneumoniae]	64	46	561
15	7	6504	7145	[gi 1762328	[Ycr59c/Yig2 homolog [Bacillus subtilis]	64	45	642
22	11	9548	9895	[gnl pt0 d100581	[unknown [Bacillus subtilis]	64	38	348
22	10	122501	21174	[gi 289760	[cme ORF1 [Bacillus subtilis]	64	44	672
26	7	14375	14199	[gi 409286	[bmrU [Bacillus subtilis]	64	30	177
27	2	1510	1134	[gi 40795	[Dx1 methylase [Desulfovibrio vulgaris]	64	51	177
29	2	614	297	[gi 2326168	[type VII collagen [Mus musculus]	64	50	318
35	2	368	721	[pir JC1151 JC11	[hypothetical 20 kK protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	64	50	354
40	1	3	449	[gi 46970	[lep10 gene product [Staphylococcus epidermidis]	64	41	447
40	7	4683	4976	[gnl pt0 e325792	[A3000005] glucose kinase [Bacillus megaterium]	64	45	294
45	7	8068	6920	[gnl pt0 d102036	[subunit of ADP-glucose pyrophosphorylase [Bacillus stearothermophilus]	64	40	1149
51	2	301	1059	[gi 43985	[nifs-like gene [Lactobacillus delbrueckii]	64	54	759
51	11	115251	18197	[gi 2293760	[AF008220] DNA-polymerase III alpha-chain [Bacillus subtilis]	64	46	3147
53	3	1157	555	[gi 1574292	[hypothetical [Haemophilus influenzae]	64	47	603
58	2	4236	1606	[gi 1573826	[alanine]-tRNA synthetase (alaS) [Haemophilus influenzae]	64	51	2631
66	1	3	1259	[gi 895749	[putative cellobiose phosphotransferase enzyme II' [Bacillus subtilis]	64	42	1257
68	5	5213	6556	[gi 436965	[lmaIA] gene products [Bacillus stearothermophilus]	64	47	1344
69	6	5356	4949	[gnl pt0 d101316	[cdd [Bacillus subtilis]	64	52	408



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
74	4	6948	5038	gi 726400	L-glutamine-D-fructose-6-phosphate amidotransferase (Bacillus subtilis)	64	50	1911
75	3	1283	1465	bbs 133379	TLS-CHOP-fusion protein (CHOP-C/EBP transcription factor, TLS-nuclear RNA-binding protein) (human, myxoid liposarcoma cells, Peptide Mutant, 462 aa) (Homo sapiens)	64	57	183
81	13	14016	14231	gi 143175	methanol dehydrogenase alpha-10 subunit (Bacillus sp.)	64	35	216
81	22	21851	22090	gnl PIR d101315	YqfA (Bacillus subtilis)	64	44	240
87	11	10046	9300	gnl PIR e323505	putative Pci1 protein (Bacillus subtilis)	64	43	747
98	7	5032	5706	gnl PIR e233880	hypothetical protein (Bacillus subtilis)	64	38	675
105	1	2	1276	gi 1657501	similar to S. aureus mercury(II) reductase (Escherichia coli)	64	45	1275
113	7	5136	6410	gnl PIR d101119	Hifs (Synechocystis sp.)	64	50	1275
119	1	2	1297	gnl PIR e320520	hypothetical protein (Natronobacterium pharaonis)	64	37	1296
123	3	1125	2156	gnl PIR e253284	ORF YDL244w (Saccharomyces cerevisiae)	64	40	1032
124	5	2331	1780	gnl PIR d101884	hypothetical protein (Synechocystis sp.)	64	50	552
129	4	3467	2709	gnl PIR d101314	YqeU (Bacillus subtilis)	64	52	759
131	1	152	3	gi 13377841	unknown (Bacillus subtilis)	64	42	150
137	11	7196	7549	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS111) - Agrobacterium tumefaciens (strain P022) plasmid T1	64	50	354
139	3	3226	2651	gi 2293301	(AF008220) YtqB (Bacillus subtilis)	64	44	576
146	10	6730	5648	gi 1322245	mevalonate pyrophosphate decarboxylase (Rattus norvegicus)	64	45	1083
147	1	2	1018	gnl PIR e137033	unknown gene product (Lactobacillus leichmannii)	64	46	1017
148	11	8430	8783	gi 2130630	(AF000430) dynein-like protein (Homo sapiens)	64	28	354
156	7	4113	3612	gnl PIR d102050	transmembrane (Bacillus subtilis)	64	31	702
157	4	1299	2114	gnl PIR d100892	homologous to Gln transport system permease proteins (Bacillus subtilis)	64	43	816
162	6	5880	6362	gi 517204	ORF1, putative 42 kDa protein (Streptococcus pyogenes)	64	58	483
164	11	9707	8769	gnl PIR d100964	homologue of ferric aquibactin transport system permease protein FatD of V. anguillarum (Bacillus subtilis)	64	40	939
175	5	3906	4598	gi 534045	antiterminator (Bacillus subtilis)	64	39	693
189	10	6154	6507	gi 581307	response regulator (Lactobacillus plantarum)	64	33	354
191	4	3519	2863	gi 149520	[phosphoribosyl anthranilate isomerase (Lactococcus lactis)]	64	46	657

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% Ident	length (nt)
202	1	76	1140	[gnl P10 e293806	[O-acetylhomoserine sulphydrylase (Leptospira meyeri)	64	47	1065
224	1	234	1571	[gi 15733193	[collagenase (pric) Haemophilus influenzae]	64	42	1336
231	3	291	647	[gi 40174	[ORF X (Bacillus subtilis)	64	43	357
253	3	709	1089	[pir JC1151 JC11	[hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	64	50	381
265	1	820	2	[gi 1377832	[unknown (Bacillus subtilis)	64	31	819
297	1	1	660	[gi 1590871	[collagenase (Methanococcus jannaschii)	64	48	660
328	1	263	21	[gi 992651	[Glnp (Saccharomyces cerevisiae)	64	41	243
5	4	8730	8098	[gi 556885	[unknown (Bacillus subtilis)	63	48	633
10	6	5178	4483	[gi 1573101	[hypothetical (Haemophilus influenzae)	63	40	696
12	11	9324	9902	[gi 806536	[membrane protein (Bacillus acidopullulicus)	63	42	579
15	10	8897	9187	[gi 722339	[unknown (Acetobacter xylinum)	63	40	291
17	2	1031	309	[gnl P10 e217602	[PinU (Lactobacillus plantarum)	63	32	723
18	8	7778	6975	[gi 1377843	[unknown (Bacillus subtilis)	63	45	804
26	4	9780	7078	[gi 142440	[ATP-dependent nuclease (Bacillus subtilis)	63	46	2703
29	5	3488	4192	[gi 1377829	[unknown (Bacillus subtilis)	63	35	705
34	11	8830	7988	[gnl P10 d101198	[ONF8 (Enterococcus faecalis)	63	45	843
35	3	1187	876	[gi 722339	[unknown (Acetobacter xylinum)	63	39	312
48	15	112509	11691	[gi 1573389	[hypothetical (Haemophilus influenzae)	63	41	819
51	11	112719	12189	[gi 142450	[ahrC protein (Bacillus subtilis)	63	35	531
55	4	3979	5022	[gi 1708640	[YeeB (Bacillus subtilis)	63	41	1044
55	15	13669	14670	[gnl P10 e311502	[thioredoxine reductase (Bacillus subtilis)	63	44	1002
68	10	9242	8919	[sp P37686 Y1AY_	[HYPOTHETICAL 40.2 KD PROTEIN IN AVTA-SELB INTERGENIC REGION (P302)	63	40	324
86	7	6554	5685	[gi 1574182	[lic-1 operon protein (licD) (Haemophilus influenzae)	63	41	870
88	8	6085	5180	[gi 2098719	[putative fibrial-associated protein (Actinomyces naeslundii)	63	43	906
96	8	5858	6484	[gi 1052803	[orf191b gene product (Streptococcus pneumoniae)	63	38	627
100	1	240	1940	[gi 7171	[fucosidase (Dictyostelium discoideum)	63	36	1701

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
104	4	3063	5765	gi 144985	phosphoenolpyruvate carboxylase [Corynebacterium glutamicum]	63	46	2703
106	8	9189	8554	gi 533099	endonuclease III [Bacillus subtilis]	63	45	636
122	6	4704	4886	gnl PID d101139	transposase [Synechocystis sp.]	63	39	183
128	7	4517	5203	gnl PID d101414	orf2 [Methanobacterium thermoautotrophicum]	63	50	687
137	4	963	1547	gnl 472920	V-type Na-ATPase [Enterococcus hirae]	63	27	585
142	7	4100	4585	gnl PID e313025	hypothetical protein [Bacillus subtilis]	63	44	486
159	5	1741	2571	gnl 1787043	(AE000184) (271; This 271 aa orf is 24 pct identical (16 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SM: P09997 [Escherichia coli])	63	39	831
171	12	8803	14406	gnl PID e324918	IgA1 protease [Streptococcus sanguis]	63	48	5604
177	1	3	347	gnl 1773150	hypothetical 14 kDa protein [Escherichia coli]	63	34	345
178	2	423	917	gnl 722339	unknown [Acetobacter xylinum]	63	41	495
178	3	794	1012	gnl 1591582	cobalamin biosynthesis protein M [Methanococcus jannaschii]	63	36	219
195	1	1377	175	gnl PID e324217	ftsQ [Enterococcus hirae]	63	33	1203
234	5	1739	1527	gnl 1591582	cobalamin biosynthesis protein M [Methanococcus jannaschii]	63	36	213
249	1	81	257	gnl 1000453	TrcR [Bacillus subtilis]	63	41	177
283	1	127	1347	gnl 396486	ORF8 [Bacillus subtilis]	63	44	1221
293	3	2804	3466	gnl 722339	unknown [Acetobacter xylinum]	63	37	663
311	1	905	486	gnl 1877424	UDP-galactose 4-epimerase [Streptococcus mutans]	63	46	420
324	1	2	556	gnl 1477741	histidine periplasmic binding protein P29 [Campylobacter jejuni]	63	36	555
365	1	219	13	gnl 2252843	(AF013293) No definition line found [Arabidopsis thaliana]	63	33	207
382	1	88	378	gnl 722339	unknown [Acetobacter xylinum]	63	40	291
385	3	364	158	gnl 2252843	(AF013293) No definition line found [Arabidopsis thaliana]	63	33	207
2	1	2495	288	gnl PID e325007	penicillin-binding protein [Bacillus subtilis]	62	42	2208
3	23	23374	24231	gnl PID e254993	hypothetical protein [Bacillus subtilis]	62	35	858
6	16	14320	13193	gnl PID e349614	nifs-like protein [Mycobacterium leprae]	62	37	1128
7	8	6819	7232	gnl PID d101324	YqhY [Bacillus subtilis]	62	32	414
7	19	15466	14207	gnl PID d101604	beta ketoacyl-acyl carrier protein synthase [Synechocystis sp.]	62	43	1260

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
7	21	17155	gnl PI01e323514	putative Fabp protein [Bacillus subtilis]	62	46	927
7	24	19526	gi 1276434	beta-ketacyl-ACP synthase III [Cuphea wrightii]	62	37	1008
12	7	5904	gi 1573768	A/G-specific adenine glycosylase (mutY) [Haemophilus influenzae]	62	43	1203
12	9	8032	gi 1591587	pantothenate metabolism flavoprotein [Methanococcus jannaschii]	62	33	762
15	11	9678	pir JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	62	43	351
17	4	2609	gi 1591081	M. jannaschii predicted coding region MJ0374 [Methanococcus jannaschii]	62	43	168
17	5	3053	gi 149570	role in the expression of lactacin F, part of the laf operon [Lactobacillus sp.]	62	44	219
22	10	8627	gnl PI01d100580	similar to B. subtilis DnaM [Bacillus subtilis]	62	43	912
30	3	865	gi 231479	[AE000627] ABC transporter, ATP-binding protein (yhcG) [Helicobacter pylori]	62	43	1179
33	5	2235	gi 413976	ipa-52r gene product [Bacillus subtilis]	62	44	600
38	11	5689	gi 148231	o231 [Escherichia coli]	62	34	435
40	17	14272	gnl PI01d101904	hypothetical protein [Synechocystis sp.]	62	43	945
42	1	3	gi 1146182	putative [Bacillus subtilis]	62	41	309
44	2	1267	gi 1786952	[AE000176] o877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment YBG8_ECOLI SM: P54746 [Escherichia coli]	62	43	2739
48	12	9732	gi 662920	repressor protein [Enterococcus faecalis]	62	32	429
51	8	5664	gnl PI01e101153	StySKI methylase [Salmonella enterica]	62	44	1518
52	3	2791	gi 1183886	integral membrane protein [Bacillus subtilis]	62	41	693
55	16	14702	gnl PI01e313028	hypothetical protein [Bacillus subtilis]	62	40	959
59	6	3418	gi 2063483	unknown [Lactococcus lactis lactis]	62	32	567
63	5	4997	gi 149771	pilin gene inverting protein (PivML) [Moraxella lacunata]	62	28	189
70	14	10002	gi 992977	bpIC gene product [Bordetella pertussis]	62	45	718
71	13	18790	gi 1280135	coded for by C. elegans cDNA cm21e6; coded for by C. elegans cDNA cm01e2; similar to melibiose carrier protein (thiomethylgalactoside permease II) [Caenorhabditis elegans]	62	62	1593
71	28	32217	gnl PI01d101312	yqcG [Bacillus subtilis]	62	35	552
74	7	11666	gi 1552753	hypothetical [Escherichia coli]	62	38	1284

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Cuntig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
80	8	9370	9609	gnl PID d102002	(AB001488) FUNCTION UNKNOWN. [Bacillus subtilis]	62	46	240
97	10	9068	7041	gi 882463	[protein-N(pI)-phosphohistidine-sugar phosphotransferase (Escherichia coli)]	62	42	2028
98	4	2106	3268	gnl PID d101496	BraE (integral membrane protein) [Pseudomonas aeruginosa]	62	42	963
102	3	2823	3539	gnl PID e13010	hypothetical protein [Bacillus subtilis]	62	24	717
103	3	2795	1242	gnl PID d102049	H. influenzae hypothetical ABC transporter; P44808 (974) [Bacillus subtilis]	62	41	1554
111	2	2035	3462	gi 581297	NISP [Lactococcus lactis]	62	44	1428
112	4	3154	4080	gi 1574379	lic-1 operon protein (licA) [Haemophilus influenzae]	62	39	927
112	6	4939	5649	gi 1574381	lic-1 operon protein (licC) [Haemophilus influenzae]	62	39	711
124	3	1137	721	gi 1573024	anaerobic ribonucleoside-triphosphate reductase (nrdD) [Haemophilus influenzae]	62	45	417
124	6	3162	2329	gi 609076	leucyl aminopeptidase [Lactobacillus delbrueckii]	62	40	834
126	7	11073	7516	gnl PID d101163	ORF4 [Bacillus subtilis]	62	38	3558
129	6	4983	4540	pti S41509 S415	zinc finger protein ERF - Chilo iridescent virus	62	48	444
131	7	4510	4103	gi 1857245	unknown [Lactococcus lactis]	62	42	408
149	2	1923	2579	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	62	41	657
149	7	5360	6055	gnl PID e123508	YioS protein [Bacillus subtilis]	62	40	696
156	1	450	238	gnl PID e254644	membrane protein [Streptococcus pneumoniae]	62	40	213
156	6	3606	2935	gnl PID d102050	transmembrane [Bacillus subtilis]	62	37	672
171	2	1779	2291	gi 43941	Elll-B Sor PTS [Klebsiella pneumoniae]	62	35	513
172	2	385	723	gi 895750	putative cellobiose phosphotransferase enzyme fli [Bacillus subtilis]	62	39	339
173	3	2599	893	gi 1591732	cobalt transport ATP-binding protein O [Methanococcus jannaschii]	62	42	1707
179	2	492	1754	gi 1574071	H. influenzae predicted coding region HI1038 [Haemophilus influenzae]	62	38	1263
181	6	2856	3707	gi 1777435	LacT [Lactobacillus casei]	62	42	852
185	2	2074	311	gi 2182197	(AC000073) Y4(N [Rhizobium sp. MCR214]	62	41	1764
200	2	1061	1984	gi 450566	transmembrane protein [Bacillus subtilis]	62	37	924
202	3	2583	1473	gi 42219	p35 gene product (AA 1 - 314) [Escherichia coli]	62	41	891
210	3	1374	1565	gi 49315	(ORF) gene product [Bacillus subtilis]	62	45	192

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
211	1	3	971	gi1247402	mannose permease subunit III-Man (Escherichia coli)	62	43	969
223	2	1495	1034	gn1PID d101190	ORF2 (Streptococcus mutans)			
228	1	34	909	gi1530063	glycerol uptake facilitator (Streptococcus pneumoniae)	62	41	462
234	2	90	917	gi12293259	(AF008220) YqI (Bacillus subtilis)	62	44	876
282	5	1765	1487	gn1PID e276475	galactokinase (Arabidopsis thaliana)	62	38	828
375	1	1	159	gi11674231	(AE000052) Mycoplasma pneumoniae, hypothetical protein homolog; stellar to Swiss-Prot Accession Number P31555, from B. subtilis (Mycoplasma pneumoniae)	62	40	159
385	5	584	357	gi11573353	outer membrane integrity protein (OIA) (Haemophilus influenzae)	62	47	228
3	19	18550	19269	gi1606162	ORF_229 (Escherichia coli)	61	41	720
7	4	2725	3225	gi12114425	similar to Synchocystis sp. hypothetical protein, encoded by GenBank Accession Number D64006 (Bacillus subtilis)	61	42	501
17	6	3326	3054	gi1149569	lactacin F (Lactobacillus sp.)	61	43	273
44	3	4061	4957	gn1PID d101068	xylose repressor (Synchocystis sp.)	61	38	897
54	11	8388	7234	gn1PID d101329	YqJH (Bacillus subtilis)	61	42	1155
57	6	3974	6037	gn1PID d101316	YqK (Bacillus subtilis)	61	42	2064
58	5	7156	6565	sp P45169 POTC_	SPERMIDINE/PUTRESCINE TRANSPORT SYSTEM PERMEASE PROTEIN POTC	61	34	792
67	1	3	692	gi1537108	ORF_254 (Escherichia coli)	61	46	690
68	9	8816	7890	gi119501	PLZ12 gene product (AA 1-188) (Lupinus polyphyllus)	61	41	927
70	15	10737	12008	gi1992976	bpIF gene product (Bordetella pertussis)	61	44	1272
72	11	9759	10202	gn1PID d101833	carboxymorsperaldine decarboxylase (Synchocystis sp.)	61	36	444
76	8	7881	7003	gn1PID d100305	farnesyl diphosphate synthase (Bacillus stearothermophilus)	61	45	879
87	4	4914	3697	gi1528991	unknown (Bacillus subtilis)	61	42	1218
87	13	11211	11361	gi1178968	(AE000407) methionyl-tRNA formyltransferase (Escherichia coli)	61	44	951
91	2	711	2989	gi1537080	ribonucleoside triphosphate reductase (Escherichia coli)	61	45	2259
105	3	2711	3499	gn1PID d101651	hypothetical protein (Synchocystis sp.)	61	44	789
115	6	7968	6473	gi1895747	putative cel operon regulator (Bacillus subtilis)	61	36	1491
123	8	7181	8518	gi11209527	protein histidine kinase (Enterococcus faecalis)	61	40	1338

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
126	6	7525	6725	gi 11787043	(AE000184) (271): This 271 aaorf is 24 pct identical (116 gaps) to 265 residues of an approx. 272 aa protein YIDA_ECOLI SM: P09997 [Escherichia coli]	61	38	801
128	1	1	639	gnl pid d101328	YqiY [Bacillus subtilis]	61	41	639
139	7	4794	5054	gi 1022726	unknown [Staphylococcus haemolyticus]	61	41	261
139	9	12632	5913	gnl pid e270014	beta-galactosidase [Thermotoga thermophilus]	61	41	6720
141	1	2552	42	gi 520541	penicillin-binding proteins 1A and 1B [Bacillus subtilis]	61	42	2511
148	16	12125	11424	gi 1552743	tetrahydrodipicolinate N-succinyltransferase [Escherichia coli]	61	42	702
162	3	4112	3456	gnl pid d101829	phosphoglycerate phosphatase [Synecocystis sp.]	61	30	657
172	3	727	1077	gnl pid d102048	B. subtilis, cellobiose phosphotransferase system, cglA; P46318 (220) [Bacillus subtilis]	61	44	351
177	3	1101	1772	gnl pid d100574	unknown [Bacillus subtilis]	61	43	672
202	2	1278	2585	gi 1045831	hypothetical protein (GB:L18965.6) [Mycoplasma genitalium]	61	36	1308
224	3	2782	3144	gi 1591144	M. jannaschii predicted coding region MJ0440 [Methanococcus jannaschii]	61	30	363
225	4	3395	3766	gi 1552774	hypothetical [Escherichia coli]	61	40	372
249	2	212	802	gi 1000453	TrcR [Bacillus subtilis]	61	42	591
254	2	841	484	gnl pid d100417	ORF120 [Escherichia coli]	61	36	360
257	1	3	350	gnl pid e255315	unknown [Mycobacterium tuberculosis]	61	42	348
293	4	3971	3657	ptr JCI151 JC11	hypothetical 20.3K protein (insertion sequence IS111) - Agrobacterium tumefaciens (strain P022) plasmid T1	61	45	315
301	1	949	17	gi 2291209	(AF016424) contains similarity to acyltransferases [Caenorhabditis elegans]	61	33	933
373	1	1066	287	gi 393396	Tb-292 membrane associated protein [Trypanosoma brucei subgroup]	61	38	780
3	24	24473	24955	gi 1537093	ORF_0153b [Escherichia coli]	60	27	483
6	5	4636	5739	gi 2293258	(AF008220) YcoI [Bacillus subtilis]	60	35	1104
6	12	11936	11187	gi 293017	ORF3 (put.): putative [Lactococcus lactis]	60	44	750
17	13	6708	6484	gi 149569	Lactacin F [Lactobacillus sp.]	60	32	225
18	7	6977	5670	gi 1788140	(AE000278) o481: This 481 aaorf is 35 pct identical (19 gaps) to 309 residues of an approx. 856 aa protein NOI1_HUMAN SM: P46087 [Escherichia coli]	60	43	1308
20	15	15878	17167	gnl pid d100584	unknown [Bacillus subtilis]	60	44	1290

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
22	1	1	243	gnl PID d102050	transmembrane [Bacillus subtilis]	60	36	243
32	10	8296	8964	gi 2293275	[AF008220] YrcG [Bacillus subtilis]	60	37	669
38	15	8837	9697	gi 40023	B. subtilis genes rpmH, rnpA, 50kd, gldA and gldB [Bacillus subtilis]	60	35	861
43	6	8610	5944	gi 171787	protein kinase 1 [Saccharomyces cerevisiae]	60	36	2667
44	1	1	1269	gnl PID e235823	unknown [Schizosaccharomyces pombe]	60	44	1269
45	10	11138	110368	gi 397488	1,4-alpha-glucan branching enzyme [Bacillus subtilis]	60	43	771
48	19	15766	14378	gnl PID e205173	orf1 [Lactobacillus helveticus]	60	39	1389
48	21	16727	16951	gnl PID d102041	[AB002668] unnamed protein product [Haemophilus actinomycetemcomitans]	60	32	225
50	1	2	898	gnl PID e246537	[OKF286] protein [Pseudomonas stutzeri]	60	31	897
62	2	638	1177	gnl PID d100587	unknown [Bacillus subtilis]	60	42	540
68	4	3590	5203	gi 1533583	H. influenzae predicted coding region H10594 [Haemophilus influenzae]	60	36	1614
70	11	5781	6182	gnl PID d102014	[AB001488] SIMILAR TO YDFR GENE PRODUCT OF THIS ENTRY [YDFR_BACSU]. [Bacillus subtilis]	60	33	402
70	12	6143	8133	gnl PID e124970	hypothetical protein [Bacillus subtilis]	60	38	1791
71	8	11701	14157	gi 580866	[ipa-12d] gene product [Bacillus subtilis]	60	33	2457
74	8	12509	11664	gnl PID d101832	phosphatidate cytidyltransferase [Synecocystis sp.]	60	45	846
76	4	4116	3367	gi 2352096	orf; similar to serine/threonine protein phosphatase [Fervidobacterium islandicum]	60	39	750
80	4	7372	7665	gi 1786420	[AE000131] 186; 100 pct identical to GB: ECODINJ_6 ACCESSION: D38582 [Escherichia coli]	60	30	294
81	6	4071	4522	gi 147402	[mannose permease subunit III-Man [Escherichia coli]]	60	35	450
86	1	940	155	gi 143177	[putative [Bacillus subtilis]]	60	26	786
92	1	1	192	gi 196348	[homoserine transsuccinylase [Escherichia coli]]	60	45	192
93	14	10619	9384	gi 1788389	[AE000297] o364; This 46 aa orf is 33 pct identical (9 gaps) to 331 residues of an approx. 416 aa protein MTRC_NEIGO SW. P43505 [Escherichia coli]	60	27	1236
94	5	5548	8121	gnl PID e129895	[AJ000496] cyclic nucleotide-gated channel beta subunit [Rattus norvegicus]	60	50	2574
97	7	5396	4533	gi 1591396	[transketolase' [Methanococcus jannaschii]]	60	43	864
102	2	2081	2833	gnl PID e120929	[hypothetical protein [Mycobacterium tuberculosis]]	60	43	753



TABLE 2  
S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
106	9	9773	9183	[gnl PID ej34782	YleN protein (Bacillus subtilis)	60	31	591
113	8	6361	6837	[gi 466875	[nifU; B1496_C1_157 (Mycobacterium leprae)	60	43	477
115	2	2755	524	[gnl PID ej28143	(AJ000337) Glucosidase II (Homo sapiens)	60	32	2232
122	7	4763	5068	[gnl PID di01876	transposase (Synechocystis sp.)	60	39	306
127	8	4510	5283	[gi 777938	Pgm (Treponema pallidum)	60	38	774
138	4	3082	2672	[gnl PID ej25196	hypothetical protein (Bacillus subtilis)	60	36	411
139	3	177	4	[gnl PID di00680	ORF (Thermus thermophilus)	60	39	174
139	11	14520	13009	[gi 537145	ORF_437 (Escherichia coli)	60	30	1512
140	2	2592	1249	[gi 1209527	protein histidine kinase (Enterococcus faecalis)	60	37	1344
141	1	210	1049	[gi 461181	E5 ORF from bp 3842 to 4081; putative (Human papillomavirus type 33)	60	34	840
141	5	5368	6405	[gi 145362	tyrosine-sensitive DAP synthase (aroF) (Escherichia coli)	60	41	1038
142	6	3558	4049	[gi 600711	putative (Bacillus subtilis)	60	37	492
148	10	7742	8713	[gnl PID ej33022	hypothetical protein (Bacillus subtilis)	60	27	972
153	5	3667	4278	[gi 2293322	(AF008220) branch-chain amino acid transporter (Bacillus subtilis)	60	42	612
155	1	1413	748	[gi 2104504	putative UDP-glucose dehydrogenase (Escherichia coli)	60	40	666
158	3	3116	2472	[gnl PID di00872	a negative regulator of pho regulon (Pseudomonas aeruginosa)	60	37	645
159	3	778	1386	[gnl PID ej308090	product highly similar to Bacillus anthracis CepA protein (Bacillus subtilis)	60	48	609
163	7	8049	8468	[gnl PID di01313	YqeN (Bacillus subtilis)	60	38	420
170	3	4170	2688	[gi 1574179	H. influenzae predicted coding region Hfl244 (Haemophilus influenzae)	60	39	1443
171	7	4717	5901	[gi 606076	ORF_0384 (Escherichia coli)	60	44	1185
183	3	2440	2135	[gi 1877427	repressor (Streptococcus pyogenes phage T12)	60	38	306
191	10	9444	8428	[gi 415664	metabolite control protein (Bacillus megaterium)	60	42	1017
200	1	139	1083	[gi 438462	transmembrane protein (Bacillus subtilis)	60	37	945
201	3	3895	1928	[gi 475112	enzyme Habc (Pediococcus pentosaceus)	60	39	1968
214	15	10910	10439	[gi 1573407	hypothetical (Haemophilus influenzae)	60	39	492
218	4	2145	2363	[gi 608520	myosin heavy chain kinase A (Dictyostelium discoideum)	60	31	219

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
226	4	2518	2351	[gi1437705]	hyaluronidase [Streptococcus pneumoniae]	60	53	168
242	1	725	3	[gi143918]	Sor regulator [Klebsiella pneumoniae]	60	41	723
245	1	1	288	[gi1304897]	EcoE type I restriction modification enzyme M subunit [Escherichia coli]	60	56	288
251	1	905	45	[gi1671632]	unknown [Staphylococcus aureus]	60	36	861
259	1	969	87	[gi1353794]	rgg [Streptococcus gordonii]	60	32	888
260	2	1492	1662	[gi131840 S318]	probable transposase - Bacillus stearothermophilus	60	26	171
274	1	836	96	[gi11592173]	N-ethylmaleine chlorohydrolase [Methanococcus jannaschii]	60	40	741
308	1	463	2	[gi11787397]	[AE000214] o157 [Escherichia coli]	60	43	462
318	1	3	308	[gnl pid e137594]	xerC recombinase [Lactobacillus leichmannii]	60	42	306
344	1	73	522	[gi1509672]	repressor protein [Bacteriophage Tuc2009]	60	32	450
5	1	576	4	[gi12293147]	[AF008220] YcM [Bacillus subtilis]	59	31	573
7	22	18140	17142	[gnl pid e280724]	unknown [Mycobacterium tuberculosis]	59	39	999
10	1	1413	4	[gi1353880]	isialidase L [Macrobactelia decoral]	59	41	1410
15	6	6463	5156	[gi1580841]	[F1] [Bacillus subtilis]	59	35	1308
22	2	479	1393	[gi142469]	lals operon regulatory protein [Bacillus subtilis]	59	34	915
22	5	2698	4614	[gnl pid e280623]	PCPA [Streptococcus pneumoniae]	59	44	1917
30	1	208	558	[gnl pid e233868]	hypothetical protein [Bacillus subtilis]	59	37	351
30	4	3678	2455	[gnl pid e202290]	unknown [Lactobacillus sakei]	59	33	1224
35	13	12201	11071	[gnl pid e238664]	hypothetical protein [Bacillus subtilis]	59	35	1131
35	14	13288	12182	[gi1457667]	Cap8M [Staphylococcus aureus]	59	39	1107
36	18	18076	17897	[gi1500535]	[M. jannaschii] predicted coding region MJ1635 [Methanococcus jannaschii]	59	33	180
38	12	6172	7137	[gi12293239]	[AF008220] YcM [Bacillus subtilis]	59	34	966
42	3	1952	3361	[gi1684845]	[pinin] [Canis familiaris]	59	40	1410
50	3	2678	1728	[gnl pid d101329]	[YcM] [Bacillus subtilis]	59	41	951
56	5	1870	2388	[gnl pid e137594]	xerC recombinase [Lactobacillus leichmannii]	59	41	519
61	6	6812	5628	[gnl pin e11516]	amlnutransferase [Bacillus subtilis]	59	40	1185
67	5	2382	1023	[gi1146190]	[2-keto-3-deoxy-6-phosphogluconate aldolase [Bacillus subtilis]	59	36	642

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
69	10	8567	8899	gi 1573628	anthranilate kinase (coaA) [Haemophilus influenzae]	59	38	333
87	12	11183	10055	gnl PID e323504	putative Fmu protein [Bacillus subtilis]	59	44	1329
113	14	11327	15894	gi 1673731	[AE000010] Mycoplasma pneumoniae, fructose-permease 118C component; similar to Swiss-Prot Accession Number P20966, from E. coli [Mycoplasma pneumoniae]	59	43	1968
115	8	8766	8521	gi 1590886	[M. jannaschii predicted coding region MJ0110 [Methanococcus jannaschii]	59	38	246
119	2	1966	1526	gnl PID e209005	homologous to ORF2 in nrDEF operons of E. coli and S. typhimurium [Lactococcus lactis]	59	43	441
128	17	11348	13118	gnl PID e279632	unknown [Mycobacterium tuberculosis]	59	38	261
140	22	121903	23388	gi 482922	protein with homology to pail repressor of B. subtilis [Lactobacillus delbrueckii]	59	40	516
148	13	9697	9014	gnl PID d102005	[AB001488] FUNCTION UNKNOWN, SIMILAR PRODUCT IN H. INFLUENZAE AND SYNECHOCYSTIS, [Bacillus subtilis]	59	32	684
149	10	7213	8244	gi 710422	[omp-binding-factor 1 [Staphylococcus aureus]	59	40	1032
164	9	6993	6013	gnl PID d100965	ferric anguibactin-binding protein precursor FatB of V. anguillarum [Bacillus subtilis]	59	41	981
164	12	8836	7823	gnl PID d100964	homologue of ferric anguibactin transport system permease protein FatC of V. anguillarum [Bacillus subtilis]	59	35	1014
177	2	401	1072	gi 289759	coded for by C. elegans cDNA CE2G3 [GenBank:214728]; putative [Caenorhabditis elegans]	59	40	672
177	7	3841	4200	gi 2313445	[AE000531] H. pylori predicted coding region HP0342 [Helicobacter pylori]	59	38	360
183	4	2768	2508	gi 509672	repressor protein [Bacteriophage Tuc2009]	59	50	261
186	6	3398	2820	gi 606080	ORF_0290; Geneplot suggests frameshift linking to o267, not found [Escherichia coli]	59	38	579
190	3	3120	1711	gi 1613768	histidine protein kinase [Streptococcus pneumoniae]	59	32	1410
194	2	1621	1019	gnl PID d100579	unknown [Bacillus subtilis]	59	40	603
198	7	5205	4306	gnl PID e313073	hypothetical protein [Bacillus subtilis]	59	38	900
220	5	4362	3958	gnl PID d101322	YqjL [Bacillus subtilis]	59	46	405
242	3	1573	2367	gi 1787045	[AE000184] {308; This 308 aa ort is 15 pct identical (35 gaps) to 305 residues of an approx. 296 aa protein PFLC-ECOLI SW: P32675 [Escherichia coli]	59	42	795
247	2	1154	1480	gi 40073	ORF107 [Bacillus subtilis]	59	39	327

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
256	1	868	2	gnl PID d101924	hemolysin [Synecocystis sp.]	59	39	867
258	1	65	820	gi 2246532	ORF 73, contains large complex repeat CR 73 (Kaposi's sarcoma-associated herpesvirus)	59	20	756
270	1	386	1126	gnl PID d102092	YfnB [Bacillus subtilis]	59	40	741
281	1	552	166	gi 666062	putative [Lactococcus lactis]	59	31	387
309	1	3	479	gi 405879	YefM [Escherichia coli]	59	38	477
363	1	2	1894	gi 915208	gastric mucin [Sus scrofa]	59	31	1893
387	2	425	84	gi 160671	S antigen precursor [Plasmodium falciparum]	59	44	342
5	6	11223	10465	gnl PID d101812	LumQ [Synecocystis sp.]	58	29	759
29	4	2098	3513	gnl PID d100479	Na <sup>+</sup> -ATPase subunit J [Enterococcus hirae]	58	39	1416
30	5	4058	3651	gi 39478	ATP binding protein of transport ATPases [Bacillus firmus]	58	34	408
33	6	2983	2210	gnl PID d101164	unknown [Bacillus subtilis]	58	45	774
36	8	5316	6179	gi 518679	orf [Bacillus subtilis]	58	32	864
43	5	5926	3971	gi 1788150	[AE000278] protease II [Escherichia coli]	58	37	1956
46	5	3704	5221	gnl PID d267329	unknown [Bacillus subtilis]	58	42	1518
48	14	11722	11066	gnl PID d101771	thiamin biosynthetic bifunctional enzyme [Synecocystis sp.]	58	34	657
52	1	1229	3	gnl PID d101291	reductase [Pseudomonas aeruginosa]	58	35	1227
53	2	702	412	gi 2333357	[AE000545] cytochrome c biogenesis protein (ccda) [Helicobacter pylori]	58	25	291
58	4	6586	5498	gi 147329	transport protein [Escherichia coli]	58	41	1089
69	5	4934	3807	gnl PID e311492	unknown [Bacillus subtilis]	58	41	1128
71	27	31357	32777	gi 2408014	hypothetical protein [Schizosaccharomyces pombe]	58	33	921
72	4	3586	2882	gi 18694	nodulin-21 (AA 1-201) [Glycine max]	58	34	705
74	3	4917	4230	gi 2293252	[AE008220] YfmO [Bacillus subtilis]	58	33	708
79	4	4594	3422	gi 1217989	ORF3 [Streptococcus pneumoniae]	58	44	1173
82	8	10585	8171	gi 882711	lexonuclease V alpha-subunit [Escherichia coli]	58	38	2415
86	17	16017	15337	gi 47642	[5-dehydroquinase hydrolyase (3 dehydroquinase) (Salmonella typhi)]	58	32	681
97	2	931	560	gi 153794	lrgg [Streptococcus gordonii]	58	32	372

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
108	2	358	2724	gi 537020	wacB gene product [Escherichia coli]	58	37	2367
111	5	4593	5240	gi 1592142	ABC transporter, probable ATP-binding subunit [Methanococcus jannaschii]	58	36	648
120	3	4421	5110	gnl PID d101320	YggX [Bacillus subtilis]	58	47	690
1	16	13131	12673	gi 662919	ONF U [Enterococcus hirae]	58	42	459
132	3	6174	4939	gi 1800301	macrolide-efflux determinant [Streptococcus pneumoniae]	58	35	1236
133	1	111	890	gnl PID e269488	Unknown [Bacillus subtilis]	58	36	780
160	11	8615	9865	gi 471901	ONF1 [Lactococcus lactis]	58	39	1251
161	6	6268	6849	gnl PID d101024	DJ-1 protein [Homo sapiens]	58	32	582
169	1	214	2	gnl PID d100447	translation elongation factor-3 [Chlorella virus]	58	31	213
187	1	487	2	gi 475114	regulatory protein [Pediococcus pentosaceus]	58	38	486
187	6	4184	4620	gi 167475	deacetylase-related protein [Craterostigma plantagineum]	58	55	237
190	2	1464	1640	gnl PID e246727	competence pheromone [Streptococcus gordonii]	58	38	177
192	2	2012	1344	gnl PID d100556	rat GCP360 [Rattus rattus]	58	44	669
206	1	1292	696	gnl PID e202579	product similar to MrBA [Lactobacillus sake]	58	35	597
216	2	2333	555	gnl PID e125036	hypothetical protein [Bacillus subtilis]	58	33	1779
217	5	5250	4321	gi 466474	cellobiose phosphorylase [Bacillus stearothermophilus]	58	38	930
217	7	5636	5106	gnl PID d102048	B. subtilis cellobiose phosphorylase system celB; P46317 (1998)	58	44	531
232	1	2	811	gi 1573777	cell division ATP-binding protein (ftsE) [Haemophilus influenzae]	58	39	810
264	1	2	715	gi 1971330	Nata [Bacillus subtilis]	58	32	714
280	1	33	767	gi 11786187	[AE000111] hypothetical 29.6 kD protein in thrC-catB intergenic region [Escherichia coli]	58	31	735
306	1	845	3	gnl PID e134780	YibL protein [Bacillus subtilis]	58	47	843
360	3	1556	1092	gp 46351 v2G0	HYPOTHETICAL 45.4 KD PROTEIN IN THIAMINASE I 5' REGION [Escherichia coli]	58	32	465
363	5	2160	1867	gi 160671	IS antigen precursor [Plasmodium falciparum]	58	51	294
372	1	806	3	gi 193194	Tb-291 membrane associated protein [Trypanosoma brucei subgroup]	58	37	804
382	2	749	519	gnl JC1151 JC11	hypothetical 20.3K protein (insertion sequence IS1131) - Agrobacterium tumefaciens (strain P022) plasmid T1	58	41	231

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
3	9	8409	7471	gi 1499745	M. jannaschii predicted coding region MJ0912 [Methanococcus jannaschii]	57	38	939
10	10	7674	7507	gi 1737169	homologue to SKP1 [Arabidopsis thaliana]	57	30	168
11	1	2	412	gnl P1D1d100139	ORF [Acetobacter pasteurianus]	57	42	411
31	4	2032	1388	gi 2293213	[AF008220] YtpR [Bacillus subtilis]	57	37	645
33	11	6931	6449	gnl P1D1e24949	hypothetical protein [Bacillus subtilis]	57	36	483
45	5	5446	5060	gi 1592204	phosphoserine phosphatase [Methanococcus jannaschii]	57	44	387
49	7	6523	7632	gi 155369	[PTS enzyme-II fructose [Xanthomonas campestris]	57	35	1110
52	6	4520	6850	gi 1574144	[single-stranded-DNA-specific exonuclease (recJ) [Haemophilus influenzae]	57	35	2331
53	5	2079	1795	gi 1843580	[replicase-associated polypeptide (coat blue dwarf virus)]	57	46	285
63	6	5312	4995	gi 2182608	[AE000094] Y4rJ [Rhizobium sp. NGR234]	57	39	318
72	15	11388	11059	gnl P1D1d100892	homologous to SwissProt: YIDA_ECOLI hypothetical protein [Bacillus subtilis]	57	40	825
79	2	2561	1815	gnl P1D1d100965	homologue of MADPH-flavin oxidoreductase Fp of V. harveyi [Bacillus subtilis]	57	44	747
82	9	9596	9763	gi 1206045	[short region of similarity to glycerophosphoryl diester phosphodiesterases [Caenorhabditis elegans]	57	35	168
86	16	115371	14493	gi 11787983	[AE000264] o288: 92 pct identical (11 gaps) to 222 residues of (fragment YD18_ECOLI SW: P28244 (223 aa) [Escherichia coli]	57	34	879
93	3	1695	1177	gi 1500003	mutator mutT protein [Methanococcus jannaschii]	57	33	519
96	6	3026	4519	gi 559882	[threonine synthase [Arabidopsis thaliana]	57	43	1494
99	14	17211	18212	gi 1773349	[BirA protein [Bacillus subtilis]	57	44	1002
112	8	7448	7903	gi 1591393	M. jannaschii predicted coding region MJ0678 [Methanococcus jannaschii]	57	30	456
113	16	18627	18328	plr A45605 A456	[mature-parasite-infected erythrocyte surface antigen HESA - plasmodium falciparum]	57	22	300
123	2	343	1110	plr F64149 F641	hypothetical protein H10355 - Haemophilus influenzae (strain Rd KW20)	57	38	768
123	4	2108	2804	gnl P1D1d102148	[AB001684] sulfate transport system permease protein [Chlorella vulgaris]	57	39	777
127	10	6477	5507	gi 1573082	[nitrogenase C (nifC) [Haemophilus influenzae]	57	35	891
128	13	9251	9790	gi 153692	[pneumolysin [Streptococcus pneumoniae]	57	38	540
131	4	2139	1363	gi 142081	[nagD gene product (AA 1-250) [Escherichia coli]	57	36	777

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
136	1	214	1221	[gb 148453]	[Spae-endocarditis immunodominant antigen (Streptococcus sobrinus, MUCOB 263, Peptide, 1566 aa) (Streptococcus sobrinus)]	57	44	1008
140	25	128701	128051	[gi 505576]	[beta-glucoside permease (Bacillus subtilis)]	57	38	1851
141	6	6395	7438	[gi 995560]	[unknown (Schizosaccharomyces pombe)]	57	41	1044
144	3	3231	2785	[gnl PID d100139]	[ORF (Acetobacter pasteurianus)]	57	42	447
155	4	5454	4564	[gi 600431]	[glycosyl transferase (Erwinia amylovora)]	57	34	891
159	9	4877	5854	[gi 290509]	[ol307 (Escherichia coli)]	57	35	978
167	11	9710	9249	[gnl PID d100139]	[ORF (Acetobacter pasteurianus)]	57	42	462
171	6	4023	4436	[gi 167402]	[mannose permease subunit III-Man (Escherichia coli)]	57	29	414
178	4	2170	1076	[gnl PID d102004]	[TAB001488] ATP-DEPENDENT RNA HELICASE DEAD HOMOLOG. (Bacillus subtilis)]	57	39	1095
190	1	145	1455	[gi 169420]	[export/processing protein (Lactococcus lactis)]	57	30	1311
198	1	298	95	[gi 522268]	[unidentified ORF22 (Bacteriophage BIL67)]	57	36	204
203	2	3195	2110	[gnl PID e281915]	[orf c0100] (Sulfolobus solfataricus)]	57	41	1086
205	1	40	507	[gi 1439527]	[Ella-man (Lactobacillus curvatus)]	57	28	468
214	7	4243	1797	[gnl PID d102049]	[in. influenzae, ribosomal protein alanine acetyltransferase, P44305 (189) (Bacillus subtilis)]	57	48	447
268	1	1767	1276	[gi 43979]	[L.curvatus small cryptic plasmid gene for rep protein (Lactobacillus curvatus)]	57	36	492
351	1	324	34	[gnl PID e275871]	[T0376. b (Caenorhabditis elegans)]	57	31	291
386	1	226	2	[gi 160671]	[S antigen precursor (Plasmodium falciparum)]	57	45	225
5	5	10486	8777	[gi 405857]	[yehu (Escherichia coli)]	56	33	1710
8	5	3674	3910	[gi 467199]	[pkac; L518_F1.2 (Mycobacterium leprae)]	56	39	237
10	3	1442	1874	[gnl PID d101907]	[sodium-coupled permease (Synechocystis sp.)]	56	36	1569
21	1	1880	333	[gi 2313949]	[AE000593] osmoprotection protein (proXA) (Helicobacter pylori)]	56	33	1548
22	129	21968	22456	[gnl PID d102001]	[TAB001488] PROBABLE ACETYLTRANSFERASE. (Bacillus subtilis)]	56	37	489
27	1	1361	1	[gi 215132]	[aa59 (525) (Bacteriophage lambda)]	56	30	1359
28	9	4667	4278	[gi 1592090]	[DNA repair protein RAD2 (Methanococcus jannaschii)]	56	29	390
33	1	3	386	[gnl PID d100139]	[ORF (Acetobacter pasteurianus)]	56	41	384

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
36	7	5122	5397	[pir PQ0051 PQ00	hypothetical protein (proC 3' region) - Pseudomonas aeruginosa (strain PAO) (fragment)	56	28	276
40	3	3137	4318	[gi 18800301	macrolide efflux determinant [Streptococcus pneumoniae]	56	27	1182
40	16	12511	13191	[gnl PID e217602	PinU [Lactobacillus plantarum]	56	38	681
48	17	13375	13323	[gi 143729	transcription activator [Bacillus subtilis]	56	35	753
75	4	1674	2594	[gnl PID d102036	membrane protein [Bacillus stearotherophilus]	56	25	921
85	3	1842	1459	[gnl PID d100139	ORF [Acetobacter pasteurianus]	56	41	384
89	7	5815	4940	[gi 853777	product similar to E coli PRA2 protein [Bacillus subtilis]	56	42	876
105	2	1160	2718	[gnl PID d101913	hypothetical protein [Synecocystis sp.]	56	37	1359
112	3	2151	3194	[gi 537201	ORF_0345 [Escherichia coli]	56	31	1044
113	4	2754	2963	[gnl PID d100140	ORF [Plum pox virus]	56	28	210
122	3	1203	2054	[gi 1649035	high-affinity periplasmic glutamine binding protein [Salmonella typhimurium]	56	30	852
124	8	1939	3694	[gnl PID e248893	unknown [Mycobacterium tuberculosis]	56	27	246
125	4	4403	4107	[gnl PID d100247	human non-muscle myosin heavy chain (Homo sapiens)	56	32	297
127	11	6608	6605	[gi 2182397	[AE000073] Y4tN [Rhizobium sp. MGR234]	56	35	204
134	5	4769	3849	[gnl PID d101870	hypothetical protein [Synecocystis sp.]	56	39	921
137	10	6814	7245	[gi 1592011	sulfate permease (cysA) [Methanococcus jannaschii]	56	34	432
142	8	5019	4582	[pit A47071 A470	orf1 immediately 5' of nifs - Bacillus subtilis	56	29	438
146	8	4676	3660	[gnl PID d101911	hypothetical protein [Synecocystis sp.]	56	32	1017
148	3	1906	2739	[gnl PID d101099	phosphate transport system permease protein psta [Synecocystis sp.]	56	36	834
150	4	4449	2743	[gnl PID e304628	probably site-specific recombinase of the resolvase family of enzymes [Bacteriophage TP21]	56	27	1707
172	3	2	208	[gi 1187791	[AE000249] f117; This 317 aa orf is 27 pct identical (16 gaps) to 301 residues of orf - approx. 320 aa protein YXXC_BACSU SW. P39140 [Escherichia coli]	56	34	207
172	7	4979	5668	[gi 396293	similar to Bacillus subtilis hypoth. 20 kDa protein, in tsr 3' region [Escherichia coli]	56	40	690
186	7	3732	3367	[gi 1732200	PFS permease for mannose subunit 1 [Phan [Vibrio furnissii]	56	36	366
187	2	2402	819	[pir SS7904 SS79	virR49 protein - Streptococcus pyogenes (strain CS101, serotype M49)	56	35	1584



TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
204	1	2772	2239	gi 606376	ORF_0162 [Escherichia coli]	56	35	534
206	2	3142	1633	gi 559861	clm [Plasmid PADII]	56	38	1710
219	3	1689	1096	gi 1146197	putative [Bacillus subtilis]	56	27	594
230	2	409	1485	pir C60328 C603	hypothetical protein 2 (sr 5' region) - Streptococcus mutans (strain ON2175, serotype I)	56	40	1077
233	4	2930	3268	gi 1041785	phoptry protein [Plasmodium yoelii]	56	24	339
273	2	1543	2724	gi 143089	lep protein [Bacillus subtilis]	56	32	1182
353	1	1	516	gnl PID j325000	hypothetical protein [Bacillus subtilis]	56	41	516
359	1	87	641	gi 1786952	(AE000176) 0877; 100 pct identical to the first 86 residues of the 100 aa hypothetical protein fragment Y8GB_ECOLI SM: P54746 [Escherichia coli]	56	46	555
363	7	4482	4198	gi 1573353	outer membrane integrity protein (colA) [Haemophilus influenzae]	56	38	285
376	1	2	508	gnl PID j325031	hypothetical protein [Bacillus subtilis]	56	33	507
18	1	836	177	gnl PID j100872	le negative regulator of pho regulon [Pseudomonas aeruginosa]	55	31	660
28	4	1824	1618	gnl PID j316518	STAT protein [Dictyostelium discoideum]	55	40	207
29	6	4496	5041	gi 1088261	unknown protein [Anabaena sp.]	55	31	546
38	16	9695	10702	gi 580905	B subtilis genes rpmI, rpmA, 50kd, gida and gldb [Bacillus subtilis]	55	31	1008
49	5	5727	6182	gi 1786951	(AE000176) heat-responsive regulatory protein [Escherichia coli]	55	29	456
51	4	2381	3241	gnl PID j101293	Ybba [Bacillus subtilis]	55	42	861
52	9	9640	10866	gi 153016	ORF 419 protein [Staphylococcus aureus]	55	23	1227
53	4	1813	1349	gi 896042	ospF [Borrelia burgdorferi]	55	30	465
60	5	4794	5756	gi 1499876	magnesium and cobalt transport protein [Methanococcus jannaschii]	55	38	963
71	9	14176	15408	gi 1857120	glycosyl transferase [Neisseria meningitidis]	55	41	1233
75	6	3189	4229	gnl PID j209890	NAD alcohol dehydrogenase [Bacillus subtilis]	55	44	1041
108	10	10488	9820	gnl PID j324997	hypothetical protein [Bacillus subtilis]	55	36	669
113	12	12273	13037	gnl PID j111496	unknown [Bacillus subtilis]	55	34	765
113	11	13007	13945	gi 1573423	1-phosphofructokinase (fruk) [Haemophilus influenzae]	55	39	939
126	5	6764	5907	gi 1750133	(AE000446) hypothetical 29.7 kD protein in bpaA-gycB intergenic region [Escherichia coli]	55	37	858

TABLE 2

S. pneumoniae Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
129	3	2719	902	gnl PID d101425	P2-peptidase [Bacillus licheniformis]	55	35	1818
138	3	2593	1610	gnl 142833	ORF2 [Bacillus subtilis]	55	37	984
140	6	6916	5633	gnl PID d100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	55	26	1284
147	3	3854	2136	gnl 472330	dihydrolipoamide dehydrogenase [Clostridium magnum]	55	39	1719
147	10	10204	8921	gnl PID e73078	dihydroorotase [Lactobacillus leichmannii]	55	38	1284
148	5	1430	4119	gnl 290572	peripheral membrane protein U [Escherichia coli]	55	29	690
148	6	4171	4650	gnl 695769	transposase [Xanthobacter autotrophicus]	55	37	480
149	14	12564	11650	gnl PID d101329	YqjG [Bacillus subtilis]	55	32	915
156	3	1113	550	gnl 231496	(AE000634) conserved hypothetical integral membrane protein (Helicobacter pylori)	55	34	564
159	10	6625	5897	gnl 290533	similar to E. coli ORF adjacent to suc operon; similar to gutR class of regulatory proteins [Escherichia coli]	55	29	729
164	3	1784	2332	gnl PID e255118	hypothetical protein [Bacillus subtilis]	55	37	549
164	5	2772	3521	gnl 40348	put. resolvase Tnp I (AA 1 - 284) [Bacillus thuringiensis]	55	35	750
164	11	7428	7216	gnl PID e249407	unknown [Mycobacterium tuberculosis]	55	38	213
167	5	3860	3345	gnl 535052	involved in protein secretion [Bacillus subtilis]	55	28	516
186	5	2880	2583	gnl 606080	ORF_0290: Geneplot suggests frameshift linking to o267, not found [Escherichia coli]	55	35	318
189	8	4311	5396	gnl PID e183450	hypothetical tcsB protein [Bacillus subtilis]	55	32	1086
192	5	3270	3079	gnl 196504	vitellogenin convertase [Aedes aegypti]	55	38	192
195	2	2454	1384	gnl 1574693	transferase, peptidoglycan synthesis (murG) [Haemophilus influenzae]	55	33	1071
198	4	3013	2471	gnl PID e313074	hypothetical protein [Bacillus subtilis]	55	29	543
214	1	373	744	gnl PID d101741	transposase [Synecocystis sp.]	55	33	372
219	4	1115	456	gnl 288301	ORF2 gene product [Bacillus megaterium]	55	30	660
263	7	3742	3443	gnl 18137	cgr-4 product [Chlamydomonas reinhardtii]	55	48	300
285	1	2	829	gnl PID d100974	unknown [Bacillus subtilis]	55	40	828
286	1	650	249	gnl 958444	ORF 118 kDa [Vibrio cholerae]	55	31	402
297	2	1229	1696	gnl 150848	pilC [Porphyromonas gingivalis]	55	39	468

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
309	2	218	982	gi 1574491	hypothetical haemophilus influenzae	55	35	765
328	2	646	224	gi 571500	prohibitin (Saccharomyces cerevisiae)	55	27	423
330	1	1140	474	gi 396397	soxS (Escherichia coli)	55	29	867
364	1	2538	1546	gi 391394	Tb-291 membrane associated protein (Trypanosoma brucei subgroup)	55	36	993
368	3	941	105	gi 160671	S antigen precursor (Plasmodium falciparum)	55	40	837
3	5	4604	3624	gi 2293176	signal transduction protein kinase (Bacillus subtilis)	54	26	981
9	11	7746	7246	gi 1146245	putative (Bacillus subtilis)	54	38	501
38	24	16213	11937	gi 1480429	putative transcriptional regulator (Bacillus stearothermophilus)	54	27	1725
40	8	5076	4882	gi 39989	methionyl-tRNA synthetase (Bacillus stearothermophilus)	54	35	195
43	4	3980	2367	gnl P16148611	ABC transporter (Lactobacillus helveticus)	54	25	1614
52	10	10843	12103	gi 1762962	FemA (Staphylococcus simulans)	54	29	1260
57	1	3	512	gi 558177	endo-1,4-beta-xylanase (Cellulomonas fimi)	54	36	510
58	1	4749	4246	gnl P16148611	hypothetical (Bacillus subtilis)	54	29	504
71	7	10684	11703	gi 510255	orf3 (Escherichia coli)	54	31	1020
71	20	127546	12737	gi 202543	serotonin receptor (Rattus norvegicus)	54	31	192
72	2	844	1098	gi 148613	arnB gene product (Plasmid F)	54	37	255
72	7	7438	6695	gi 1196496	recombinase (Moraxella bovis)	54	38	744
74	10	14043	13465	gi 1200342	ORF 3 gene product (Bradyrhizobium japonicum)	54	32	579
74	12	16483	15995	gi 2317798	maturase-related protein (Pseudomonas alcaligenes)	54	30	489
86	3	2877	2155	gi 46988	orf9.6 possibly encodes the O unit polymerase (Selenomonas enterica)	54	34	723
89	5	4433	3921	gi 147211	phnO protein (Escherichia coli)	54	41	513
90	1	3	464	gi 2317798	maturase-related protein (Pseudomonas alcaligenes)	54	30	462
96	10	8058	8510	gnl PID102015	AB001488) SIMILAR TO SALMONELLA TYPHIMURUM SLYV GENE REQUIRED FOR SURVIVAL IN MACROPHAGE (Bacillus subtilis)	54	32	453
97	6	4662	3604	gi 1591394	transketolase (Methanococcus jannaschii)	54	30	1059
106	11	10406	12010	gi 606286	ORF_0637 (Escherichia coli)	54	32	1605
147	8	8663	7404	gnl PID101615	ORF_ID:031987; similar to (SwissProt Accession Number P33401) (Escherichia coli)	54	35	1260

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
171	4	2477	3223	[gi 1439528	ELIC-man (Lactobacillus curvatus)	54	36	747
174	2	2068	1787	[gnl PID d100518	motor protein (Homo sapiens)	54	35	282
188	1	526	1188	[gnl PID a250352	unknown (Mycobacterium tuberculosis)	54	31	663
198	5	3582	2884	[gnl PID e11074	hypothetical protein (Bacillus subtilis)	54	33	699
207	1	1	1641	[gnl PID d101813	hypothetical protein (Synecocystis sp.)	54	24	1641
210	1	2	655	[gi 2293206	(AF008220) Ymp (Bacillus subtilis)	54	29	654
225	2	966	2357	[gnl PID e130194	RII16.1 (Caenorhabditis elegans)	54	39	1392
241	1	1681	347	[gnl PID d101813	hypothetical protein (Synecocystis sp.)	54	26	1335
263	2	907	1395	[gnl PID d101886	transposase (Synecocystis sp.)	54	30	489
263	6	3450	2979	[gi 160671	S antigen precursor (Plasmodium falciparum)	54	47	474
277	3	2517	1363	[gi 1196926	unknown protein (Streptococcus mutans)	54	30	1155
307	1	828	4	[gi 2293198	(AF008220) YtGP (Bacillus subtilis)	54	28	825
325	1	19	768	[gi 2182507	(AE000083) Y41H (Rhizobium sp. MGR234)	54	37	750
332	2	898	590	[gi 1591815	ADP-ribosylglycohydrolase (drag) (Methanococcus jannaschii)	54	32	309
385	4	240	479	[gi 530878	amino acid feature: N-glycosylation sites, aa 41...43, 46...48, 51...53, 72...74, 107...109, 128...130, 132...134, 158...160, 163...165; amino acid feature: Rod protein domain, aa 169...340; amino acid feature: globular protein domain	54	49	240
7	25	19702	119493	[gnl PID a255111	hypothetical protein (Bacillus subtilis)	53	32	210
73	1	2497	2013	[gnl PID d102015	(AB001488) SIMILAR TO SALMONELLA TYPHIMURUM SLVY GENE REQUIRED FOR SURVIVAL IN MACROPHAGE (Bacillus subtilis)	53	25	465
29	11	9042	10121	[gi 143331	alkaline phosphatase regulatory protein (Bacillus subtilis)	53	31	1080
33	1	1479	1009	[pi S10655 S106	hypothetical protein X - Pyrococcus woesei (fragment)	53	33	471
36	6	4583	5134	[gnl PID e116029	unknown (Mycobacterium tuberculosis)	53	30	552
38	14	8521	8898	[gi 580904	homologous to E coli rnpA (Bacillus subtilis)	53	30	378
52	7	7007	8686	[gi 1377811	unknown (Bacillus subtilis)	53	29	1680
54	17	17555	19564	[gi 1666069	orf2 gene product (Lactobacillus leichmannii)	53	36	2010
56	1	1	681	[gi 1592266	restriction modification system S subunit (Methanococcus jannaschii)	53	32	681

## TABLE 2

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	length (nt)
57	10	9431	8887	gi 1788543	[AE000310] f351: Residues 1-121 are 100 pct identical to YOJL_ECOLI SW: P3394 (122 aa) and aa 152-351 are 100 pct identical to YOJL_ECOLI SW: P3394 [Escherichia coli]	53	31	945
61	1	429	4	gnl PIU e233667	B0024.12 [Caenorhabditis elegans]	53	33	426
71	1	5772	4	gi 393394	TD-291 membrane associated protein [Trypanosoma brucei subgroup]	53	33	5769
72	3	894	2840	gi 2293378	[AF008220] YtaD [Bacillus subtilis]	53	27	1947
73	14	9793	9212	gi 1778556	putative cobalamin synthesis protein [Escherichia coli]	53	32	582
88	7	5217	4342	gi 2098719	putative fibrillar-associated protein [Actinomyces naeslundii]	53	38	876
93	5	2195	1688	gi 563366	gluconate oxidoreductase [Gluconobacter oxydans]	53	33	708
96	9	6632	7762	gi 517204	[ORF], putative 42 kDa protein [Streptococcus pyogenes]	53	42	1131
108	8	7629	8600	gi 149581	maturation protein [Lactobacillus paracasei]	53	32	972
128	9	6412	6972	gnl PID e317237	unknown [Mycobacterium tuberculosis]	53	36	561
128	112	8429	9253	gi 3311070	pentraxin fusion protein [Xenopus laevis]	53	31	825
148	1	3	950	pir A61607 A616	probable hemolysin precursor - Streptococcus agalactiae (strain 74-160)	53	36	948
163	2	2162	3022	gi 1755150	nocturnin [Xenopus laevis]	53	30	861
171	3	2304	2624	gi 1732200	PTS peptidase for mannose subunit (IPMan [Vibrio furnissii])	53	32	321
182	5	3785	3051	gnl PID d100572	unknown [Bacillus subtilis]	53	35	735
209	1	2948	1915	gi 1778505	lactic enterobactin transport protein [Escherichia coli]	53	28	1014
218	5	3884	2406	gi 40162	murE gene product [Bacillus subtilis]	53	34	1479
250	1	473	790	gnl PID e334726	YibH protein [Bacillus subtilis]	53	30	318
275	3	1	1611	gnl PID d101314	Yqew [Bacillus subtilis]	53	35	1611
332	1	544	2	gi 409286	bcrU [Bacillus subtilis]	53	31	543
2	2	2543	3445	gnl PID e233879	hypothetical protein [Bacillus subtilis]	52	39	903
3	122	22402	23376	gi 38969	lacf gene product [Agrobacterium radiobacter]	52	36	975
5	3	8094	2356	gnl PID e324915	IgaI protease [Streptococcus sanguis]	52	32	5739
22	26	19961	20212	gi 152901	[ORF 3] [Spirochaeta aurantia]	52	35	252
22	31	23140	24666	gi 289262	[comE ORF] [Bacillus subtilis]	52	32	1527
27	6	5397	4801	gi 39573	P20 (AA 1-178) [Bacillus licheniformis]	52	35	597

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
35	10	8604	7357	gi 508281	putative O-antigen transporter [Escherichia coli]	52	27	1248
45	4	4801	3662	gnl PID d102243	(AB005541) homologs are found in E. coli and H. influenzae; see SWISS-PROT ACC# P42100 [Bacillus subtilis]	52	36	1140
48	16	14385	13726	gnl PID e205174	orf2 [Lactobacillus helveticus]	52	25	660
49	4	5321	5755	gi 2317740	(AF013987) nitrogen regulatory IIA protein [Vibrio cholerae]	52	19	435
54	4	2773	4668	gi 1500472	M. jannaschii predicted coding region MJ1577 [Methanococcus jannaschii]	52	36	1896
54	6	5250	4969	gi 2182453	(AF000079) Y410 [Rhizobium sp. NGR234]	52	40	282
66	6	8400	6955	gi 443140	TrkG protein [Escherichia coli]	52	30	1446
71	26	10659	11112	gnl PID e31499	unknown [Mycobacterium tuberculosis]	52	23	654
75	2	1673	1035	gnl PID d102271	(AB001683) FcA [Streptomyces sp.]	52	27	639
81	3	1439	2893	gnl PID e311458	rhamnose kinase [Bacillus subtilis]	52	32	1455
81	8	4987	5781	gi 147403	mannose permease subunit II-P-Man [Escherichia coli]	52	37	795
83	21	20687	21853	gi 143165	phosphoribosyl aminimidazole carboxylase II (Pur-K; tsg start codon) [Bacillus subtilis]	52	37	1167
86	6	5785	4592	gi 1276879	EpsF [Streptococcus thermophilus]	52	26	1194
86	20	19190	17861	gi 454844	ORF 3 [Schistosoma mansoni]	52	26	1530
96	13	10540	9659	gi 288299	ORF1 gene product [Bacillus megaterium]	52	33	882
111	1	2	2026	gi 148309	Cytolysin B transport protein [Enterococcus faecalis]	52	27	2025
112	2	1457	2167	gi 471234	orf1 [Haemophilus influenzae]	52	33	711
118	3	2931	2365	bbs 151233	Hip-24 kDa macrophage infectivity potentiator protein [Legionella pneumophila, Philadelphia-1, Peptide, 184 aa] [Legionella pneumophila]	52	33	567
122	9	5646	5951	gi 8214	myosin heavy chain [Drosophila melanogaster]	52	36	306
122	11	6159	6374	gi 434025	dihydrolypamide acetyltransferase [Pseudomonas carbinolicus]	52	52	216
134	6	4880	6313	gi 151733	M protein trans-acting positive regulator [Streptococcus pyogenes]	52	43	1434
135	3	1238	2716	gnl PID e245024	unknown [Mycobacterium tuberculosis]	52	35	1479
141	3	1681	2319	gnl PID d100573	unknown [Bacillus subtilis]	52	32	639
161	4	2562	5624	gi 1146243	22-4S identity with Escherichia coli RNA-damage inducible protein putative [Bacillus subtilis]	52	36	2463
173	2	968	183	gi 121569	putative orf: GT9_orf434 [Mycoplasma pneumoniae]	52	30	786

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
198	6	4400	3567	[gnl pid e313010]	hypothetical protein [Bacillus subtilis]	52	26	834
210	12	8844	9107	[gi 497647]	DNA gyrase subunit B [Mycoplasma genitalium]	52	38	264
214	10	5264	5431	[gi 550697]	envelope protein [human immunodeficiency virus type 1]	52	36	168
225	1	15	884	[gi 1552773]	hypothetical [Escherichia coli]	52	34	870
230	1	39	362	[gnl pid d100582]	unknown [Bacillus subtilis]	52	28	324
287	1	871	2	[gnl pid e335028]	protease/peptidase [Mycobacterium leprae]	52	29	870
363	2	1305	4	[gi 393394]	17B-291 membrane associated protein (Typhosoma brucei subgroup)	52	32	1302
23	2	2048	1173	[gnl pid e254943]	unknown [Mycobacterium tuberculosis]	51	30	876
29	3	742	1521	[gi 929900]	5'-methylthioadenosine phosphorylase [Sulfolobus solfataricus]	51	31	780
45	1	410	1597	[gi 1877429]	integrase [Streptococcus pyogenes phage T12]	51	32	1188
48	26	19227	18946	[gi 2314455]	[A6000633] transcriptional regulator (tena) [Helicobacter pylori]	51	33	282
73	5	4276	4016	[gi 474177]	alpha-D-1,4-glucosidase [Staphylococcus xylosum]	51	31	261
81	11	8935	12057	[gi 311070]	pentraxin fusion protein [Xenopus laevis]	51	31	3123
83	5	1195	1986	[gnl pid d101316]	[vqfi [Bacillus subtilis]	51	33	792
98	10	7531	8538	[gi 41500]	[ORF 3] (AA 1-352); 38 kD (put. fixX) [Escherichia coli]	51	28	1008
113	6	3908	5173	[gi 466882]	[pps1; B1496_C2_189 [Mycobacterium leprae]	51	27	1266
124	1	326	57	[gi 2191168]	[A6007270] contains similarity to myosin heavy chain [Arabidopsis thaliana]	51	32	270
129	10	7286	6816	[gi 1046241]	[orf14 [Bacteriophage HP1]	51	30	471
143	3	4963	3983	[gi 1354935]	[probable copper-transporting atpase [Escherichia coli]	51	26	981
148	15	11359	10226	[gi 2293256]	[A6008220] putative hippurate hydrolase [Bacillus subtilis]	51	36	1134
149	8	6003	7313	[gi 1633572]	[herpesvirus saimiri ORF7] homolog [Kaposi's sarcoma-associated herpes-like virus]	51	21	1311
151	9	112092	11550	[gnl pid e281580]	hypothetical 40.7 kD protein [Bacillus subtilis]	51	34	543
159	6	2555	3208	[gi 146944]	[CMP-H-acetylneuraminic acid synthetase [Escherichia coli]	51	36	654
174	1	1797	4	[gi 1773166]	[probable copper-transporting atpase [Escherichia coli]	51	28	1794
265	4	2231	1773	[gnl pid e256400]	[anti-P falciparum antigenic polypeptide [Saimiri sciureus]	51	18	459
277	2	643	1311	[pur 532915 5329]	[p1D protein - Neisseria gonorrhoeae]	51	33	669

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	Match accession	Match gene name	% sim	% ident	Length (nt)
350	1	890	3	gi 290509	po307 [Escherichia coli]	51	30	888
363	4	1228	4485	gi 1707247	partial CDS (Caenothabditis elegans)	51	23	3258
367	1	1701	4	gi 393394	TD-29) membrane associated protein (Trypanosoma brucei subgroup)	51	32	1698
15	5	5174	4497	gnl pid e58151	F3 [Bacillus subtilis]	50	38	678
16	4	2220	2582	gnl pid e125010	hypothetical protein [Bacillus subtilis]	50	29	363
19	5	2591	4159	gi 1552733	similar to voltage-gated chloride channel protein [Escherichia coli]	50	30	1569
25	4	2701	1997	gi 887849	ORF_1219 [Escherichia coli]	50	27	705
35	1	211	417	gnl pid e236697	unknown [Saccharomyces cerevisiae]	50	33	207
39	4	3416	5152	gnl pid d100974	unknown [Bacillus subtilis]	50	27	1737
51	7	4000	5181	gi 1592027	carbamoyl-phosphate synthase, pyrimidine-specific, large subunit [Methanococcus jannaschii]	50	27	1182
51	9	7179	8303	gi 1591847	type I restriction-modification enzyme, S subunit [Methanococcus jannaschii]	50	28	1125
52	8	8740	9534	gi 144297	acetyl esterase (XynC) [Calducellum saccharolyticum]	50	34	795
52	16	16591	15770	gi 2108229	basic surface protein [Lactobacillus fermentum]	50	34	822
57	7	6031	6316	gi 2275264	60S ribosomal protein L7B [Schizosaccharomyces pombe]	50	40	306
71	21	29348	28383	gnl pid d101328	YqjA [Bacillus subtilis]	50	30	966
86	12	11155	10769	gnl pid e324964	hypothetical protein [Bacillus subtilis]	50	24	387
93	2	1205	330	gi 1066016	similar to Escherichia coli pyruvate, water dikinase, Swiss-Prot Accession Number P23538 [Pyrococcus furiosus]	50	24	876
96	5	1673	2959	gnl pid e322433	gamma-glutamylcysteine synthetase [Brassica juncea]	50	29	1287
98	2	218	1171	gi 151110	leucine-, isoleucine-, and valine-binding protein [Pseudomonas aeruginosa]	50	30	954
103	4	3303	2785	gi 154330	O-antigen ligase [Salmonella typhimurium]	50	31	519
115	5	6480	5980	gi 895747	putative cel operon regulator [Bacillus subtilis]	50	26	501
129	11	7559	7305	gi 1216475	skeletal muscle ryanodine receptor [Homo sapiens]	50	32	255
129	13	8192	7965	gi 152271	319-kDa protein [Rhizobium meliloti]	50	30	228
151	5	7634	6819	gi 40348	put. resolvase Tip I (AA 1 - 284) [Bacillus thuringiensis]	50	35	816
153	1	1	597	gnl pid d102015	(AB001488) SIMILAR TO NITROREDUCTASE [Bacillus subtilis]	50	29	597



TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
155	5	5986	5432	gi1276880	EpsG (Streptococcus thermophilus)	50	28	555
160	9	7390	6123	gi1786981	(AE000179) o331: 92 pct identical to the 133 aa hypothetical protein YBHE_ECOLI SW: P52697; 26 pct identical (7 gaps) to 167 residues of the 373 aa protein MLE_TRAICU SW: P46057; SW: P52697 (Escherichia coli)	50	30	1068
163	6	7196	8091	gn1P10 d101313	YgeN (Bacillus subtilis)	50	22	696
167	6	5232	3940	gi1413926	Ipa-2r gene product (Bacillus subtilis)	50	27	1293
169	2	807	130	gn1P10 e304540	endolysin (Bacteriophage Bastille)	50	35	678
171	5	3168	4025	gi1606080	ORF_0290; Geneplot suggests frameshift linking to o267, not found (Escherichia coli)	50	27	858
210	11	8151	8414	gi1330038	HRV 2 polyprotein (Human rhinovirus)	50	25	264
164	1	1538	135	gi1393396	Tb-282 membrane associated protein (Trypanosoma brucei subgroup)	50	31	1404
10	7	5911	5090	gi1144859	ORF B (Clostridium perfringens)	49	24	822
26	5	10754	9788	gi1142440	ATP-dependent nuclease (Bacillus subtilis)	49	31	987
66	7	9177	8398	gi1414170	trkA gene product (Methanobacillus marisnigellii)	49	26	1380
77	6	5164	4648	gn1P10 e285122	RexX protein (Mycobacterium smegmatis)	49	28	717
82	13	13289	13249	gn1P10 e255091	hypothetical protein (Bacillus subtilis)	49	20	561
91	9	4866	4531	gi140067	X gene product (Bacillus sphaericus)	49	26	336
112	5	4019	4948	gi11574380	lhc-1 operon protein (lhcB) (Haemophilus influenzae)	49	27	930
129	7	6058	4949	gn1P10 e267587	Unknown (Bacillus subtilis)	49	35	1110
135	5	3875	4438	gi139573	P20 (AA 1-178) (Bacillus licheniformis)	49	25	564
154	2	1423	1953	gn1P10 d101102	regulatory components of sensory transduction system (Synecocystis sp.)	49	29	531
156	5	2878	1637	gn1P10 d101732	hypothetical protein (Synecocystis sp.)	49	25	1242
173	5	3500	2940	gi1490324	ORF X gene product (unidentified)	49	30	561
182	1	1057	2	gi1331002	first methionine codon in the ECLF1 ORF (Saimiriine herpesvirus 2)	49	25	1056
192	6	5152	3667	gi12394472	(AF024499) contains similarity to homeobox domains (Caenorhabditis elegans)	49	23	1686
253	4	1129	1350	gi1511116	SRM4 protein (Saccharomyces cerevisiae)	49	23	222
277	1	600	136	gi1396844	ORF (18 kDa) (Vibrio cholerae)	49	32	465
327	3	1415	887	gi1733524	phosphatidylinositol-4,5-diphosphate 3-kinase (Dictyostellium discoideum)	49	24	549

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
365	1	1436	132	[gi 393394	TB-291 membrane associated protein (Trypanosoma brucei subgroup)	49	31	1305
33	7	4461	3277	[gi 145644	codes for a protein of unknown function (Escherichia coli)	48	26	1185
40	2	652	1776	[gnl PID e290649	ornithine decarboxylase (Nicotiana tabacum)	48	29	1125
67	4	1377	2184	[gi 1772652	2-keto-3-deoxygluconate kinase (Haloterrax alicantel)	48	30	1008
74	2	4269	3871	[gi 2182678	[AE000101] Y4WJ (Rhizobium sp. NGR231)	48	27	399
81	2	1326	541	[gi 153672	lactose repressor (Streptococcus mutans)	48	33	786
81	3	2981	3646	[gi 146042	fuculose-1-phosphate aldolase (fucA) (Escherichia coli)	48	30	666
97	1	602	51	[gi 153794	egg (Streptococcus gordonii)	48	29	552
110	1	1	3132	[gi 1381114	prtb gene product (Lactobacillus delbrueckii)	48	23	3132
131	5	2914	2147	[gnl PID e183811	Acyl-ACP thioesterase (Brassica napus)	48	27	768
133	3	3494	2628	[gnl PID e261988	putative ORF (Bacillus subtilis)	48	27	867
139	6	4231	4599	[gi 1049388	[ZK470.1 gene product (Caenorhabditis elegans)	48	23	369
139	8	5036	5665	[gi 1022725	unknown (Staphylococcus haemolyticus)	48	29	630
140	12	11936	11007	[gnl PID d102049	H. Influenzae, ribosomal protein alanine acetyltransferase; P44305 (189) (Bacillus subtilis)	48	27	930
146	9	5670	4654	[gi 1591731	malvalonate kinase (Methanococcus jannaschii)	48	24	1017
161	3	1280	2374	[gnl PID d101578	Collagenase precursor (EC 3.4. - -) (Escherichia coli)	48	24	1095
172	11	110581	11048	[gnl PID d101132	hypothetical protein (Synecocystis sp.)	48	27	468
182	4	2930	2586	[gi 40067	X gene product (Bacillus sphaericus)	48	37	345
210	15	110786	11196	[sp PI1940 LE29	[LATE EMBRYOGENESIS ABUNDANT PROTEIN D 29 (LEA D-29)	48	30	411
214	12	6231	6482	[gi 40389	non-toxic components (Clostridium botulinum)	48	26	252
221	1	704	3	[gi 1573364	H. Influenzae predicted coding region H1032 (Haemophilus influenzae)	48	27	702
227	2	647	3928	[gi 1673693	[AE000005] Mycoplasma pneumoniae, C09.orf118 Protein (Mycoplasma pneumoniae)	48	30	3282
251	2	480	758	[gnl PID e236697	unknown (Saccharomyces cerevisiae)	48	31	279
363	3	1874	1122	[gi 18137	cgrc 4 product (Chlamydomonas reinhardtii)	48	40	753
389	1	505	2	[gi 18137	cgrc-4 product (Chlamydomonas reinhardtii)	48	38	504
3	21	20879	22258	[gnl PID e264778	putative maltose-binding protein (Streptomyces coelicolor)	47	33	1380

TABLE 2  
S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
6	4	4089	4658	gi 39573	p20 (AA 1-178) [Bacillus licheniformis]	47	23	570
15	1	3736	3760	gnl PID100572	unknown [Bacillus subtilis]	47	25	1977
35	15	14516	13263	gi 1773351	Cap5L [Staphylococcus aureus]	47	20	1254
51	6	3547	4002	pir A37024A370	32K antigen precursor - Mycobacterium tuberculosis	47	38	456
55	8	10154	9273	gi 39848	103 [Bacillus subtilis]	47	26	882
92	4	1753	3276	gnl PIDe280611	PCPC [Streptococcus pneumoniae]	47	35	1524
127	9	5589	5386	gi 1786458	IAE000134 (120; This 120 aaorf is 76 pct identical (0 gaps) to 42 residues of an approx. 48 aa protein Y127_HAEIN SW: P43949 [Escherichia coli])	47	32	204
110	2	1232	1759	gnl PIDe266555	unknown [Mycobacterium tuberculosis]	47	23	528
140	4	4951	3542	gnl PIDd100964	homologue of hypothetical protein in a rapamycin synthesis gene cluster of Streptomyces hygroscopicus [Bacillus subtilis]	47	24	1410
151	3	6814	6200	gi 1522674	M. jaanschi predicted coding region MJECL4 [Methanococcus jaanschi]	47	27	615
157	3	803	1174	gnl PIDd101320	YggZ [Bacillus subtilis]	47	25	372
178	5	1267	2155	gi 2367190	IAE000390 0334; sequence change joins ORFs YGJR & YGJS from earlier version (YGJR_ECOLI SW: P42599 and YGJS_ECOLI SW: P42600) [Escherichia coli]	47	30	1113
273	1	2	1549	gnl PIDe254973	autolysin sensor kinase [Bacillus subtilis]	47	32	1548
100	2	880	644	gi 1835755	zinc finger protein Png-1 [Mus musculus]	47	22	217
54	14	14182	12638	pir S43609 S436	rofa protein - Streptococcus pyogenes	46	24	1345
88	1	2	1018	gnl PIDe223891	xylose repressor [Anaerococcus thermophilum]	46	27	1017
96	7	4553	5860	gnl PIDd101652	ORF_ID:034745; similar to [SwissProt Accession Number P45272] [Escherichia coli]	46	23	1308
112	1	1127	3	gi 2209215	IAF0043251 putative oligosaccharide repeat unit transporter [Streptococcus pneumoniae]	46	24	1125
122	13	7308	7982	gi 1054776	hrr44 gene product [Homo sapiens]	46	34	675
127	14	9198	8125	gi 146286	latA gene product [Actinobacillus pleuropneumoniae]	46	28	1074
137	4	7093	6197	gi 153794	199 [Streptococcus gordonii]	46	26	897
140	8	8220	7723	gi 1235795	pullulanase [Thermoanaerobacterium thermosulfurigenes]	46	21	498
140	9	9205	8315	gi 407878	leucine rich protein [Streptococcus equisimilis]	46	27	891

TABLE 2

S. pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
162	1	1	1125	gi11143209	[ORF7: Method: conceptual translation supplied by author (Shigella sonnei)]	46	25	1125
199	1	1	585	gi11947171	[AF000299] No definition line found (Caenorhabditis elegans)	46	28	585
223	1	1971	1477	sp102562 MYSS	[MYOSIN HEAVY CHAIN, SKELETAL MUSCLE (FRAGMENTS)]	46	27	495
232	2	760	1608	gi11016112	[ycf18 gene product (Cyanophora paradoxa)]	46	28	849
292	1	687	2240	gi11673744	[AE000011] Mycoplasma pneumoniae, xylidine deaminase; similar to GenBank Accession Number C53112, from M. litium (Mycoplasma pneumoniae)]	46	29	468
30	8	5843	6472	gi11788049	[AE000270] o235; This 235 aa ORF is 29 pct identical (10 gaps) to 198 residues of an approx. 216 aa protein YTXB_BACSU SW: P06368 (Escherichia coli)]	45	24	630
48	6	1461	3868	gi1722339	[unknown (Acetobacter xylinum)]	45	29	408
60	1	107	2	gi11699079	[coded for by C. elegans cDNA yk41h4.3; coded for by C. elegans cDNA yk148g10.5; coded for by C. elegans cDNA yk152g5.5; coded for by C. elegans cDNA yk59a10.5; coded for by C. elegans cDNA yk41h4.5; coded for by C. elegans cDNA cm20g10; coded]	45	36	306
72	16	14371	14874	gi11321500	[NADH dehydrogenase (ubiquinone) (Artemia franciscana)]	45	25	504
99	7	9158	7941	gi1152192	[mutation causes a succinoglucon-minus phenotype; ExoB is a transmembrane protein; third gene of the exoYQ operon; putative (Rhizobium meliloti)]	45	28	1218
127	12	7046	6606	bhs1153689	[Hit-B10 utilization protein (Haemophilus influenzae, type b; DL42, HTHI TN106, Peptide, 506 aa) (Haemophilus influenzae)]	45	24	441
137	5	1561	2619	gi1472921	[v-type Na-ATPase (Enterococcus hirae)]	45	33	1059
209	1	774	364	gi1304141	[restriction endonuclease beta subunit (Bacillus coagulans)]	45	28	411
314	1	604	2	gi11480457	[latex allergen (Hevea brasiliensis)]	45	31	603
20	18	19782	20288	gi1433942	[ORF (Lactococcus lactis)]	44	26	507
87	8	7030	6452	gi1537207	[ORF_f277 (Escherichia coli)]	44	26	579
166	5	4909	4037	gn11010308082	[membrane transport protein (Bacillus subtilis)]	44	25	873
247	1	818	75	gn11010100718	[ORF1 (Bacillus sp.)]	44	20	744
32	3	1885	3876	gi12351768	[PspA (Streptococcus pneumoniae)]	43	24	1992
36	17	15467	18256	gi11045719	[M. genitalium predicted coding region MG064 (Mycoplasma genitalium)]	43	26	2790
54	15	14656	17343	gi1520541	[penicillin-binding proteins 1A and 1B (Bacillus subtilis)]	43	27	2688
67	2	696	1352	gi1516934	[yja gene product (Escherichia coli)]	43	29	657
139	2	2416	338	gi1396400	[similar to eukaryotic Na+/H+ exchangers (Escherichia coli)]	43	24	2079

TABLE 2

S pneumoniae - Putative coding regions of novel proteins similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)	match accession	match gene name	% sim	% ident	length (nt)
298	1	3	809	[gi 413972	lipa-48r gene product [Bacillus subtilis]	43	24	807
387	1	47	427	[gi 2313652	[AF016669] No definition line found [Caenorhabditis elegans]	43	30	381
185	4	4221	3127	[gi 2182399	[AE000073] Y46P [Rhizobium sp. NCR234]	41	25	1095
340	1	582	70	[gnl pfid 218681	CDP-diacylglycerol synthetase [Arabidopsis thaliana]	41	20	513
363	6	4205	1914	[gi 1256742	R27-2 protein [Trypanosoma cruzi]	41	27	2292
368	2	2	943	[gi 21783	LMW glutenin (AA 1-356) [Triticum aestivum]	41	34	942
155	3	4489	2361	[gi 42023	member of ATP-dependent transport family, very similar to mdr proteins and hemolysin B, export protein [Escherichia coli]	40	18	1629
365	2	95	1438	[gi 1633572	Herpesvirus saimiri ORF7 homolog [Kaposi's sarcoma-associated herpes-like virus]	40	21	1344
1	3	2979	3860	[gnl pfid 101908	hypothetical protein [Synecocystis sp.]	39	26	882
1	5	3814	4647	[gnl pfid 101961	hypothetical protein [Synecocystis sp.]	39	19	834
26	6	14035	10724	[gi 142439	ATP-dependent nuclease [Bacillus subtilis]	38	20	3312
47	1	3	4916	[gi 612549	INF-180 [Petromyzon marinus]	36	23	4914

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
1	4	1428	3009
1	6	4611	4964
3	2	818	994
3	3	1182	1574
3	7	5182	6497
3	25	25046	25396
3	26	25625	26317
6	2	1519	1689
6	14	12875	12618
6	15	13215	12841
6	18	15977	15390
7	12	9955	9419
7	13	10161	9910
8	6	3915	4280
9	9	6024	5704
10	8	6909	6298
10	9	7116	6888
10	11	7968	7672
12	1	1140	4
12	3	1779	1456
14	2	1913	1434
16	1	1	241
16	5	5675	3087
17	1	324	34
17	3	1451	1050
17	9	4890	4465
20	14	14544	115893

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
21	3	3359	2589
21	5	4802	4482
22	21	17099	17362
22	25	19467	19982
22	33	25540	25764
22	35	26388	26218
22	36	26382	27572
23	7	6655	6032
23	8	7132	6653
24	1	36	518
25	5	3009	2641
27	4	4819	4223
27	5	4789	4956
28	5	3017	1797
28	8	4272	3850
28	10	5028	4597
28	11	5746	5072
29	7	5596	4919
29	8	5039	5518
29	9	5595	8207
30	9	6511	6263
31	6	2664	2344
32	5	5203	5538
33	8	5327	4668
34	10	8024	7740
34	12	9360	8641
34	13	9667	9377

**TABLE 3**  
*S. pneumoniae* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
34	18	11104	11902
35	11	9688	8588
35	12	11071	9670
36	2	334	1041
36	12	11120	10893
36	13	10993	11388
36	15	12172	14595
38	7	4269	4577
38	8	4480	5001
38	10	5517	5711
38	17	10732	11376
40	3	1728	3143
43	1	172	5
43	7	8884	8732
43	8	9568	9071
44	4	4831	6831
45	3	1204	3665
46	4	3875	3468
46	7	6074	7081
48	5	3196	3582
48	8	4579	4229
48	11	9323	8922
48	16	13042	12494
48	20	16342	15764
48	24	17971	18351
48	30	21979	21776
49	1	209	3



TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
50	4	3307	2672
51	5	3239	3598
52	11	12146	12883
54	7	5588	5187
54	8	6013	5459
54	9	6004	6210
54	16	17685	17506
55	9	110515	110123
55	12	11947	12141
56	3	935	1387
56	4	1496	1939
57	3	1624	2130
57	4	2100	2501
58	6	7541	7335
59	1	2	430
59	4	2416	2736
59	5	2734	3063
59	8	4743	5549
59	9	5459	5929
60	6	5741	6451
61	3	2395	1772
61	5	3316	3176
64	1	2722	2
66	2	1180	3147
66	8	9082	9495
67	3	1343	1182
69	2	1165	980

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
70	5	4059	3922
70	6	4215	4057
70	9	5268	5504
71	15	20351	21901
71	16	21859	22338
71	19	26204	27556
72	9	8458	8081
73	4	1815	4216
73	6	4214	4582
73	7	4369	4773
73	10	7183	6428
73	15	9462	9668
76	1	524	195
76	2	867	535
76	11	8602	9210
80	6	7924	8109
81	1	244	2
81	10	6631	8931
83	4	1872	1150
83	17	16810	16460
84	3	4464	2929
86	2	2147	1092
86	4	3606	2875
86	19	16767	17114
87	5	5326	5000
87	7	6459	6001
87	9	7224	7006

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
87	18	17930	17670
87	19	18275	17928
88	2	1619	1840
88	4	2711	2878
88	9	6252	6016
89	3	2634	1621
89	9	7371	6868
90	2	899	2395
90	3	1143	952
91	3	2959	3141
91	4	3170	3691
91	6	4253	4573
93	1	391	2
93	6	2648	2379
93	8	4533	3712
96	1	3	182
96	2	904	632
96	3	1407	1147
96	4	1250	1420
97	9	7043	6753
99	15	18522	18692
99	17	19717	19541
100	2	4094	1980
103	1	48	299
103	6	4924	4373
104	5	6142	6735
105	7	6098	6517

**TABLE 3** *S. pneumoniae* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
106	1	1	363
106	10	9832	10212
108	1	2	268
111	3	3417	3788
111	4	3809	4606
115	10	10854	10438
116	3	2873	2121
118	2	2274	1357
122	4	2698	2333
122	10	5858	6199
122	12	6301	7416
124	2	346	890
128	4	2544	3368
129	1	689	102
129	2	1011	724
129	8	6454	6056
129	9	6540	6277
129	12	7809	7621
131	3	1433	756
131	10	5972	5673
134	11	11838	11209
135	2	625	1140
136	4	2913	3830
137	2	325	134
139	12	14027	14521
139	13	14840	14532
139	14	15363	14875

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
140	20	19822	20838
142	1	1	285
146	3	760	479
146	4	1149	778
146	7	3604	2885
146	13	8223	9401
146	14	9399	10676
146	15	10052	9750
147	7	7488	7276
147	9	8913	8647
148	7	5298	4765
149	1	2	1936
149	3	2557	2880
149	9	6258	6070
150	2	1355	579
150	3	2556	1909
153	3	2061	2842
154	3	1953	1741
155	2	2181	1411
156	8	4550	4311
157	1	37	294
159	2	631	780
159	4	1384	1722
159	7	3271	4017
161	2	1332	1018
165	3	5535	4945
166	6	5406	4972

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
167	9	6075	6395
169	5	2828	3205
170	7	6485	6243
170	8	6964	6362
170	9	7303	6962
170	11	8790	7906
171	9	7150	7476
172	5	2298	1948
173	4	2913	2677
175	2	659	835
175	3	893	1789
176	2	1487	546
176	3	2200	1466
177	9	4686	4925
177	10	4923	5177
177	11	5111	5347
177	13	7396	8703
178	6	3452	3724
181	5	1853	2473
182	2	2112	1102
182	3	2617	2006
183	2	2126	2320
185	5	4683	4219
185	6	4846	4634
187	4	2940	3557
188	4	3686	4363
188	5	4183	4821

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
188	6	5882	6493
189	5	3143	2844
189	9	5956	5564
191	1	618	4
191	11	10357	10001
192	3	2861	2268
192	4	3081	2878
192	7	6800	5331
193	3	997	839
194	4	2315	2127
195	5	6249	4543
195	6	6620	6231
196	2	1553	1849
197	1	1	861
198	9	6844	6644
200	5	5329	5769
200	6	5993	6595
204	5	1914	3276
205	2	447	1709
209	4	2038	2460
209	5	2458	2682
210	10	7370	8230
210	13	9029	10441
210	14	10439	10705
214	5	2581	2330
214	9	5065	5277
214	11	5996	5754

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
217	2	541	194
218	2	914	1432
218	3	1430	1972
218	6	3639	3821
219	1	458	39
220	1	869	600
223	4	2617	1964
227	1	1	510
234	4	1539	1312
234	6	2116	1838
235	1	52	312
235	2	310	687
238	1	660	64
246	1	1	270
248	1	3	362
248	2	443	1222
254	3	2789	792
258	2	1179	1616
260	3	1770	2123
263	1	653	177
263	4	2244	1900
263	5	3569	2973
266	1	1	342
266	2	177	1022
270	2	1124	1681
272	1	857	186
275	2	1684	2295



**TABLE 3** *S. pneumoniae* - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
278	1	2	406
282	1	714	191
282	4	1463	1134
287	2	1119	826
288	1	540	4
289	1	684	4
291	5	1589	1858
293	2	2539	2925
294	1	21	608
296	2	494	700
296	3	670	843
302	1	261	530
309	3	559	350
310	2	249	1889
316	2	2087	1818
317	2	1048	584
318	2	313	777
319	3	477	133
327	2	912	607
331	1	1	549
333	1	2	535
333	2	465	82
333	3	127	342
341	1	1	705
345	2	895	701
346	2	750	199
349	1	1	198

TABLE 3

S. pneumoniae - Putative coding regions of novel proteins not similar to known proteins

Contig ID	ORF ID	Start (nt)	Stop (nt)
350	2	81	413
355	1	44	973
358	2	636	448
360	2	948	628
364	2	1639	1265
378	1	345	1004
379	2	683	510
381	1	109	693
385	1	150	4
385	2	269	30

148

## (1) GENERAL INFORMATION:

(i) APPLICANT: Charles Kunsch  
Gil H. Choi  
Patrick S. Dillon  
Craig A. Rosen  
Steven C. Barash  
Michael R. Fannon  
Brian A. Dougherty

(ii) TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

(iii) NUMBER OF SEQUENCES: 391

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
(E) COUNTRY: USA  
(F) ZIP: 20850

## (v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
(D) SOFTWARE: ASCII Text

## (vi) CURRENT APPLICATION DATA:

149

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Brookes, A. Anders

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PB340P1

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8512

150

## (2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5625 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCAAGCAAAA CCAGCTACAG CTAAAGGAAC TTACGTAACA AACTTGACTA TCACAACTAC	60
TCAAGGTGTT GGTATCAAAG TTGACGTAAA CTCACCTTAA TCAGTAGTTA AAGTAATGTA	120
AAAAAGTTGA AGACGCTATG TCTCACTTT TTTTGATGTA CGACGGCCAT CTTGTATAGT	180
AGATGTGTAC TATTCTAGTT TCAATCTACT ATAGTAGCTC AGAAGTCGGT ACTTAAACGT	240
GCTATATCAA AACCAGTCCT TGAAAAACGT GGACTGGTTT CGTGTTTGGA TTATTACCTT	300
GAACGACATG CGTTAAAAAGT TAGTTGAACC GCCGTATGCC GAACGGACGT ACGGTGGTGT	360
GAGAGGGGCT AGAGATTATC CCCTACTCGA TTTGGAATC TAGTGGAAAT AATCTGGAAT	420
AGTCCATCGA GCTTTCTAAT ACTCTTCGAA AATCTCTTCA AACCACGTCA ACGTCGCCTT	480
GCCGTGCGTA TGGTTACTGA CTTTCGTCAGT TCTATCCACA ACCTCAAAAC AGTGTTTTGA	540
GCTGACTACG TCAGTTCCAT CTACAACCTC AAAACAGTGT TTTGAGCAAC CTGCGGCTAG	600
TTTCCTAGTT TGCTCTTTGG TTTTCATTGA GTATAACACA TTGTTAGAAG TTGCTTTAAA	660
TTTCCTAATC AGTTTGTTCA CATTTACCTT CGATATATTA TATCCCATAG TTAAGGTTGG	720
TCATACAGAT GATTATAGTC ATGGAGCCGT AAAACTTAGT GTTCTTTTAG TTGACAAAGA	780
TGCCATGAAA AAAATATTG TAACTGTAAT AGGATATTTT GAAATAAATA TAGATGAAAA	840
TATCACCGAT ATTCTATACG TAAATGGTAC TGCTATTCTT TATCTTTATT TACGTTCAAT	900
TGTTTCAATA GTTTCGGCAA TTGATAGCAG TGAAGCAATG TTGCTACCTA TCATTAATGT	960
TTTAGAGTTA CTAGATAAAT CTCAACCTTT TGAAGAAGAA TAATTTATTA GCTCACTAAA	1020
TTGAGGGTAA GAAAAAGTAA AAGCAGTAAG AAAAATGTCT TGCATTATAC AGCAACCTTT	1080
TGGGAATGAG TGGATGGATT GAATAAAATT TGATTAAGAG TGGATGATTT ATCTGTAGAT	1140
TATTATTGGA CAGTTAGTCT TGAAGTAGTC TAAGAATTAG GTTATAATCA GTAGAAGCCT	1200
TGCTAATAAT GAGGAGGTTA GTTTATGTAT AGTAGACTGA ATCTAAAATA GTACGAAACA	1260
ATTGCTAAAA CATTTATAGA AATTAATTTT ACTTTCCCAA TCGATTTGTT CTCATCTTAT	1320
TTCAATCCGC TATATATTAT GGTATCGAAT CTTATCAGA ATGATAAAAT TAATCAATTG	1380
ATATCTGATT ACAAACAGAA TATGAAAGCT TTTTATATCA CTATTGAAAA ATTTATACGA	1440

151

GATGATGAAA GCCTTAAGTG TTATTTTATA AAGGTTATTT CAAGTCGTTT CAAGGTAACA	1500
AGTCTAGATC AGATTGAAGC TGATAAAACG ATACAAAGAA AATATTCAAG TGAGCTAAAA	1560
AAATTTATTG GATTTTATAA TGAGATTATT TGTGAGGAAA ATAGTTTCCT ACATGTACGA	1620
AAGAGGTGGT CGAGTTGGTT TAGGTAGTCG ATGCGTGAGT TGATAATTCT CAGGGTATGG	1680
ACTTCTTTT CATGAATGAG GTAAAAGAGC AGGTATTGTT TAGAGACAAT CATTCTGAGC	1740
ATATTTTCTG GATAGAGGGA GTATCCGATT TTATGATGAA AGTTAATACC GCCCTCTGGT	1800
GAGAAGATGA GTAGGTTGGT AATTTAACT ATTAAACAGA ATTTTGTATT AAAAGTATTA	1860
TTTCATGAGA GAAATCCTAA TTTCACAATC CATAGGCAAA CGCTTGCAAT TCGTTTTTTA	1920
TTGGACTATA ATAGGTTGGT ATAAAGCCTT CTGTAGTAAT AAAATGTAGA AGGTGTAGAA	1980
AGTAAGGATT TAGAATATTT GTAGTTAAAA ACACAATGTT GCTATTCCTT ACGATAGGGA	2040
GATAGATATG GCAATGATAG AAGTGAACA TCTTCAGAAA AATTTTGTGA AGACTGTTAA	2100
GGAACCGGGC TTGAAGGGGG CTTTGGCGTC CTTTATTCAT CCTGAAAAGC AGACCTTTGA	2160
AGCGGTCAAG GATTTGACCT TTGAGGTTC AAAAGGGCAG ATTTTAGGAT TTATCGGGGC	2220
AAATGGTGCT GGGAACTCGA CAACCATTA AATGCTGACA GGAATTTTGA AACCAACATC	2280
TGGTTTTTGT CGGATTAACG GCAAGATTCC CCAGGACAAT CGGCAAGATT ATGTCAAAGA	2340
TATTGGCGTA GTCTTTGGAC AACGCACCCA GCTATGGTGG GATTTGGCTC TGCAAGAGAC	2400
CTACACTGTC TTAAAAGAGA TTTATGATGT GCCAGACTCG CTCTTTCATA AGCGTATGGA	2460
CTTTTTGAAT GAAGTCTTGG ATTTGAAGGA CTTTATCAAG GATCCCGTGC GGACTCTTTC	2520
ACTGGGACAA CGGATCGGGG CGGATATTGC GGCCTCCTTG CTCCACAATC CCAAGGTTCT	2580
TTTTTTAGAT GAGCCGACCA TTGGTTTGGA CGTTTCGGTT AAGGATAATA TTCGTCGGGC	2640
AATTACTCAG ATCAATCAAG AGGAAGAAAC TACCATTCTT TTGACCACTC ACGATTTGAG	2700
TGATATTGAG CAACTTTGTG ATCGGATTTT CATGATTGAC AAGGGGCAAG AGATTTTGA	2760
TGGAACGGTG AGCCAACTCA AGGAGACCTT TGTAAGATG AAGACTCTCT CTTTGAACCT	2820
GCTACCAGGT CAAAGTCATC TCGTCTCTCA CTATGACGGT CTGTCTGATA TGACCATTGA	2880
TAGACAAGGA AACAGCCTCA ACATTGAATT TGATAGTTCT CGCTACCAGT CAGCTGACAT	2940
TATCAAGCAA ACCCTGTCTG ATTTTGAAAT CCGCGATTTG AAGATGGTGG ATACGGATAT	3000
TGAGGATATT ATCCGTCGCT TCTACCGAAA GGAGCTCTAG GATGATCAAA TTGTGGAGAC	3060
GTTATAAACC CTTTATCAAT GCAGGGGTTT AGGAGTTGAT TACTTACCGA GTCAACTTTA	3120
TTCTCTATCG GATTGGCGAT GTCATGGGG CTTTGTGGC CTTTATCTC TCGAAGGCTG	3180

152

TCTTTGATTC TTCGCAAGAG TCTTTGATTC AGGGCTTCAG TATGGCGGAT ATCACCTCT	3240
ACATCATCAT GAGTTTTGTG ACCAATCTTC TGA TAGATC CGATTCGTCC TTTATGATTG	3300
GGGAGGAGGT CAAGGATGGC TCCATTATCA TGCCTTTGTT GCGACCAGTG CATTTTGCGG	3360
CCTCCTATCT TTTACCGAG CTTGGTTCCA AGTGGTTGAT TTTATCAGC GTTGGCCTTC	3420
CATTTTAAAG TGTCATTGTC TTGATGAAAA TCATATCGGG TCAAGGTATT GTAGAGGTGC	3480
TAGGATTAAC TGTCATTTAT CTTTTTAGCT TAACGCTCGC CTATCTGATT AACTTTTTCT	3540
TTAATATTTG CTTTGGATTT TCAGCCTTTG TGTTTAAAAA TCTTTGGGGT TCCAACCTAC	3600
TTAAGACTTC CATAGTGGCT TTTATGTCGG GGAGTTTGAT TCCCTTGCCA TTTTTTCCAA	3660
AGGTTGTTTC AGATATCTTC TCCTTTTTCG CTTTTTCATC CTTGATTTAT ACTCCAGTTA	3720
TGATCATTGT TGGAAAATAC GATGCCAGTC AGATTCCTCA GGCACCTCCT TTGCAGTTCT	3780
TCTGGCTCTT AGTGATGGTG GGATTGTCCT AGTTAATTTG GAAACGGGTC CAGTCCTTTA	3840
TCACCATTCA AGGAGGTTAG TATGAAAAA TATCAACGAA TGCATCTGAT TTTATCAGA	3900
CAATACATCA AACAAATCAT GGAATATAAG GTAGATTTTG TGGTTGGTGT CTTGGGAGTC	3960
TTTCTGACTC AAGGCTTGAA TCTCTTGTTC CTCAATGTCA TCTTTCAACA TATTCCATTC	4020
CTAGAAGGCT GGACCTTTCA AGAGATAGCT TTCATTTATG GATTTTCCTT GATTCCTAAG	4080
GGAAATGGACC ATCTCTTTTT TGACAATCTC TGGGCACTAG GGCAACGCCT AGTCCGAAAA	4140
GGGGAGTTTG ACAAGTATCT GACTCGTCCC ATCAATCCTC TCTTTCACAT CCTAGTTGAA	4200
ACCTTTCAGA TTGATGCCTT GGGTGAATC TTAGTCGGTG GTATTTTATT GGGAACAACA	4260
GTGACCAGCA TTGTTTGGAC TCTTCCAAAA TTCCTGCTTT TCCTAGTTTG TATTCTTTT	4320
GCGACCTTGA TTTATACTTC TCTTAAAATC GCAACAGCCA GTATCGCCTT TTGGACTAAG	4380
CAGTCAGGCG CCATGATTTA CATCTTCTAT ATGTTCAATG ACTTTGCTAA GTATCCGATT	4440
TCTATTTACA ATTCTCTTCT TCGTTGGTTG ATTAGCTTTA TCGTGCCTTT CGCCTTTACA	4500
GCCTACTATC CAGCTAGCTA TTTCTTACAG GAAAAGGATG TGTCTTTTAA CGTAGGAGGT	4560
TTGATGTTGA TTTCTCTGGT TTTCTTTGTT ATTTCCCTTA AACTTTGGGA TAAGGGCTTA	4620
GATTCCTACG AAAGTCCGGG TTCGTAAAAG CTAAAGTAAG ACTAAAATCA AGAAAGAAAC	4680
TTATGATGTT TGTAAATTGAA GAAGTCAAGG ATGAAAATCA AAAAAAGGCA GTTGTCTGCTG	4740
AGGTTTTGAA GGATTTGCCA GAATGGTTTG GAATCCCAGA AAGCACACAA GCCTATATAG	4800
AAGGAACCAC GACACTGCAA GTTTGGACCG CCTATCAGGA GAGTGATTTG ACTAGATTTG	4860
TAAGCTTATC CTATTCGAGT GAAGATTGTG CAGAGATTGA TTGTCTCGGC GTAAAAAAGC	4920
TTATCAAGGT AGAAAAATTG GGAGCCAATT GCTTGCTACT TTAGAGAGTG AAGCTCGTAA	4980

153

AAAAGTTGCT TATCTGCAGG TCAAAACAGT GGCAGAAGGT TCTAATAAAG ATTATGATCG	5040
AACAAATGAC TTTTATCGAG GTCTTGGCTT TAAAAAGTTA GAGATTTTTC CTCAACTATG	5100
GAATCCGCAA AATCCTTGTC AGATTTTGAT TAAAAAGCTT GAATAATATT ACTTGACATC	5160
TATTCTCAGA GTGCTATACT GTAAGTGTA TCGCCGATTT AGCTTAGTTG GTAGAGCAAG	5220
GCACTCGTAA AGCCTAGGTT ATAGGTAGAT AAACGACTGA GGATTTGAAA AAATAGATAG	5280
GTAGAAGATA ACCGTTAAGC CTTACTCTTA GCGGTTATTT ATATTGTTTA ATAGCGCTAA	5340
TATTTTATCA ATTATGCCTG TTTCGTGTT TCTGGTAGTT GTTCAAGTTT ATTGCTACTA	5400
TTTTTGATGG TATGAATGTG CTTATAATGT ATCCCGGTTA ACGAAAGTTT TGGACTTATA	5460
CTCTTCGAAA ATCTCTTCAA ACCACGTCAA CGTCCGCTTG CCGTGCCTAT GGTATGACT	5520
TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGT GACTACGTCA GTTCCATCTA	5580
CAACCTCAAA ACACTGTTTT GCCCAATCTG CGGCTAGTTT CCTAG	5625

## (2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7571 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

CTCTCCAGCT TTCCTTGCGA GTTGGCCATG TTGTGTCTTT AAGAAGCTA AAAATATCTC	60
CAATAAAACG CATCGCTCTC TCCTATCTCG TTTCTCTGTG TGTAGTGAC TTGCCACAAT	120
GCTTACAAAA TTTATTTACT TCTAGTCGTG TAGGCTTGAG GTTCCGCTG ATCTTGATTG	180
AATAGTTTCT CGAACCACAA ACCGCACAAG CTAGGCTTGC TTTTTTTAGT GCCATAACGC	240
CTCCATCTTA TCCATTATAA CAAGAAAGCT AGGCTTTGAC AAGCATCTTA GCGAAATAGA	300
TTGACTATCG AATCCCATAT TGTTTGAGCC TTTTCCTTAA TCTTCGCATC TGAGATAGCC	360
CGGCTAGCCT CATCTACTAG ACTTTGCGCA CGCCCTCGAA TATCAGACAA ATTATCATCT	420
GTCTGGCTAT TATCATTGGT TTGTACTTGT CTTTTTGTAT TGGCTGGTGC AATTCCATTT	480
TGCTTATAAG CATTTTCAAC CGTAAAGGTA CTTCTGGCG TATAAGGTAA AATGGTATTG	540
GCAATGTTTC TAAAGACATG AGCTGCACCG TTTGAAGTAG AGCCAGCTAG ATAGTGGTTT	600
TCATCAGTGG TCGGAAAGCC AAGCCAGTGG CTAATCACTA CATCCGGAGT ATAACCAATT	660
ACCCACTGGT CACTTGTGTA CTCGGATTG AAAACTGCTT CAGTTGTTCC AGTTTTCCCT	720



154

GCCATGACAT AGTCTGCAGG CGATGAACTA ATACCGGTAC CGTTGGTGAA AGTCCCCAAC	790
ATCATACTGG TCATCTTGTC AGCTACAGAC TTATCAATCA CCCGTTTTTG TGAATTTTTA	840
TGACTCGCAA TAACTTGTC ACTAGCATTT TCAATTCTAC TAATAAAATG AGCTTCAGGC	900
ATTAAACCTT CATTTGCAAA GGCGGCGTAT GCTTGAGCCA TTTGAAGAGG GTTGGTTTCA	960
ACACCGCTTC CCAAGGCGAC ACCAAGAACA CGGTCGACCT TTTCCATGTT GAGTCCGAAT	1020
TTTTCGCCCTG CCTCAAAAGC CTTGTGCGACA CCCAAATCAT TAACAGTGGC AACAGCAGGT	1080
AGATTAAGCG ATTCTGCCAA GGCTTGATAC ATAGGAACTT CTCGACTCGT TTTGATCCCT	1140
GCATAGTTAT CAACCTTATA GCTGTCATAC TGCATGGTAT GGTATCCAA CTGCTTATTC	1200
AAAGCCCAGC TTGCTTCAAC TGCTGGCGTA TAAACAATA AAGGCTTAAT TGTAGAACCA	1260
GGACTACGCT TTGATTGGGT TGCATAGTTG AAATCCGGA ATCCAGTTT ATCAITGTCA	1320
GCAACTTGAC CGACAACCTC ACGAACTCCC CCTGTTTTCG GTTCGAGGGC TACACTTCCT	1380
GATTGAGCAA ACGTTCCATC CTCTGCCCTC GGAAATAGCG ATGTGTTTTC ATAAACAATC	1440
TGCATATTTG CTTGGTAGTT TTGGTCCAGC TCTGTGTAAA TGCGGTAGCC ATTATTGACA	1500
ATCTCTTCCT CTGTTAGATT ATACTTGAA ACAGCTTCAT TAACCACCGC ATCAAAATAA	1560
GAGGGGTAAC GGTAATCTGA GATTTTTCTT TCATACTTAT CGTGCAATTG CGAAGTCATA	1620
TCAACTTCAG CAGCTTTGGT TTCTTGGTTT TTATCAATAT ATCTGCTGC AACCATATTC	1680
TGCAAGACAG TATCGCGCCG ATTAGTAGAA TCTTCTACGG AATTCAAGGG ATTATACAGT	1740
TCCGGCCCCCT TGAGCATCCC TGCCAGAGTC GCAGCTTGAT CCAGACTCAC TTCTGATGCA	1800
GAAACTCCAA AGTATTTCTT ACTCGCATCT TCTACACCCC ACACACCATT TCCAAAATAA	1860
GCGTTGTTAA GGTACATGGT TAGAATTTGC TCCTTACTAT ATTTTTGCT TAATTCTAAG	1920
GCAAGGAAAA ATTCTTTTCG TTTTCTCTCA ACAGTTTGAT CCTGCGATAA ATAGGCGTTT	1980
TTAGCCAGCT GTTGGGTAAT GGTAGAGCCA CCACCTGAAC GTCCAGCAGT GACAATAGCC	2040
AAGAAAAAAC GGCCATAGTT AATCCCGTCA TTTTATAGA AAGAACGGTC TTCTGTCGCA	2100
ATAACAGCAT TCTGCAAGTT TTTACTGATG TCAGTCAGCT CAACATAGGT TCCCTTTTGA	2160
CCAGACAAGG CACCAGCCTC TTTTCTTCA CGGTCAAAAA TAAGAGTCCG AGTTTTCAAG	2220
GCATTTTGCA AATCATTGAC ATTGGTCGAC TTGGCTACAG CAAACAAATA GATTCCAAC	2280
AGCAAGCCTG CACTCAAACC TAGTATAAGG ATAATCTTG TTAGATGATA ACGACGCCAG	2340
AATTTTCGAA TCGGACCTAC TTGGGCTAAT TTTTTCGAT CACTACGAGA GCGACGTAAG	2400
ATAGTAGAAT CAGACTCCTC TAGTTCACCT GTTCTTTTTT TAAAAAGAGA AAGAAATTC	2460
TCAAATAATT TATCTAATTT CATGCGTTTA TTTTATCATC TTCATCATAG GAAGACAAGA	2520

155

ATTTAGCTAT TTCCTATCCA AATAGGGCTT TTTTGTGTAC AATATCTGTA TGCAATTCAC	2580
ATTTACATTA CCCGCTCTC TACCTCAAAT GACAGTAAAG CAATTACTTG AGGAACAACT	2640
CCTCATCCCT AGAAAAATCC GTCATTTTTT GAGAATCAAG AAACATATTT TGATAAATCA	2700
AGAAGAAGTC CACTGGAAGG AAATCGTAAA TCCTGGAGAT GTTTGCCAGT TGACTTTTGA	2760
CGAGGAAGAT TATTCCTAAA AGACGATCCC TTGGGGCAAC CCAGACTTAG TGCAGGAAGT	2820
TTATCAAGAT CAACACTTGA TTATTGTAAA CAAACCAGAG GGGATGAAAA CGCATGGTAA	2880
TCAACCAAAC GAAATTGCCC TTCTTAACCA TGTGAGTACC TATGTTGGCC AAACCTGCTA	2940
TGTCGTTTAT CGTCTGGACA TGGAAACCAG TGGCTTAGTT CTCTTTGCCA AAAATCCTTT	3000
TATCCTGCCC ATTCTCAATC GCTTATTGGA GAAAAAAGAG ATTTCTAGAG AATATTGGGC	3060
TCTAGTTGAT GGAAATATCA ACAGAAAAGA ACTTGTTTTC AGAGACAAAA TTGGACGTGA	3120
TCGCCATGAT CGTAGAAAA GAATAGTTGA TGCAAAAAAT GGGCAATATG CTGAAACGCA	3180
TGTAAGCAGA TTAAAGCAAT TCTCAAACAA GACTTCCTTG GCTCATTGCA AGCTAAAGAC	3240
AGGGCGAACC CATCAGATTC GTGTGCACCT TTCGCATCAT AATCTTCCTA TCCTGGGAGA	3300
CCCTCTCTAT AATAGTAAAT CAAAGACAAG CCGGCTTATG CTTCATGCCT TCCGACTTTC	3360
CTTTACCCAC CCACTTACTT TAGAGAAGCT AACTTTCACT ACCCTTTCAA ATACATTTGA	3420
AAAAGAATTA AAAAAGAATG GATGATCGTG TCATCCATTT TTCCATATAA AAAAGCAAGA	3480
CCACAAAGCC TTGCTTTCTA TCAACTCAAG AATTATTTAG CAATTTTTCG GAAGTATTCA	3540
AGAGTACGAA CAAGTTGTGC AGTGTATGAC ATTTGTTTGT CGTACCATGA TACAACCTTA	3600
ACCAATTGTT TACCGTCAAC GTCAAGAACT TTAGTTTGAG TTGCGTCAAA CAATGAACCG	3660
TAAGACATAC CTACGATATC TGAAGATACG ATTGGATCTT CTGTGTAACC GTATGATTCT	3720
TTTGAAGCTG CTTTCATAGC TGCCTTCACT TCATCAACAG TAACGTTCTT TTCAAGAACT	3780
GCTACCAATT CAGTAACTGA TCCAGTTGGA GTTGGAAACGC GTTGTGCAGA TCCGTCAAGT	3840
TTACCATTCA ATTCTGGGAT TACAAGACCG ATAGCTTTTG CAGCACCAGT TGAGTTAGGA	3900
ACGATGTTTG CAGCACCAGC GCGAGCACGG CGAAGGTCAC CACCACGGTG TGGTCCGTCA	3960
AGGATCATTT GGTCAACAGT GTAAGCGTGG ATAGTAGTCA TCAATCCTTC AACAACACCA	4020
AAGTTGTCTT GAAGAGCTTT AGCCATTGGA GCCAAGCAGT TTGTAGTACA TGAAGCACCT	4080
GAGATAACTG TTTCAGTACC GTCAAGAACG TCGTGGTTAG TGTGAATAC AACTGTTTTA	4140
ACGTCGTTTC CACCAGGAGC AGTGATAACA ACTTTTTTAG CTCCACCTTT AAGGTGTTTT	4200
TCAGCTGCTT CTTTCTTAGC AAAGAAACCA GTAGCTTCAA GAACGATTTT TACACCGTCA	4260

156

GTAGCCCACT CGATTTGTTT TGGATCACGT TCAGCAGAAA CTTTGAAGAA TTTACCGTTA	4320
ACTTCAAATC CACCTTCTTT AACTTCAACA GTACCGTCGA AACGACCTTG AGTTGTGTCG	4380
TATTTCAACA AGTGTGAAG CATACTGGA TCTGTAAGGT CGTTGATGCG TGTAACTTCA	4440
ACACCTTCTA CGTTTTGGAT ACGACGGAAG GCAAGACGAC CGATACGTCC GAAACCGTTA	4500
ATACCAACTT TAACTACCAT TAGTGATTTC CTCCTTATGA AAATCATGAA ATTTTATTG	4560
TGAAAAGAGT AACTTGAATC ACTACAAATC ACCTTTC AACCTATTA TACAATATT	4620
TGAGTTGAAT TGCAAGTATG GCCATTGTTT TTCTATGTTA GTTTCCTTTT AAGACTGTAA	4680
ACCAAGGAAT CCCTTACTAT TCATAGCATA ACGATTCTAT AGGATCCATT TTACTAATCT	4740
TACGCGCCGG GAAGTAGGCT GAGACATAAC CAAGTAATAG AGCGAAACT AGAGTTCCTA	4800
AAACAGATAA AAGATTTAAT TTAATAAATC TAGTGATGGA TGGGTAAAG TGACTTACAA	4860
TCGCATTGCG CAACTTCCC ACCCTTGTG CAACCAAAA TGCCAGCAGC AAGGCGATGC	4920
CTACAATCCA GATAGCCTCG TAAATAAATA TTCCTTTGAC ATCAGGATTC TGATAACCAA	4980
CTGCTTTCAT GACACCTATT TCCTTGAAC GTTGCATGAT ATTGATGTAA ATAATGATAC	5040
CAATCATAAC CGTGCTACC ACAATAGCTT GTGATGAAAG CACAATCAAT AATCCCTGAA	5100
TAACACGAAT AAAGGTAATC ACAATATCAA GAACTCTCTG TTGAGAAAGC ACAGTATACT	5160
TCTTATTTTT CTGTAATTCT TCTGTACTA CTTTTGCTG TGATGGATCT TTGAGTTCCA	5220
AGATAAATA AGATACAGCT TTCGTAAATC CAGCCTCTTT CAAAATCGTT TCCATTTGAT	5280
GAGACAGCAT GAAACTGTTG CTGCTCTCCA TGTCATCTTC ATCATTGATT ACACGTACAA	5340
TCTTCGTTG AAATTGAGCA ATCTTACTAG TTTCGGCAGC ACTTCTACA ATGCTGGCTG	5400
AGACTGATTT GCCAATAAGA TCATTAGCTG TCAAAATTTT TCCTGTCTGT TCATTCCAAT	5460
TTTTTAGTAA ACTGCTTGA ATCGTTAATC CCTGTTTATT TGTATCAGTA TAGAGGGATC	5520
CAGCCAACAC TTGTCCGTC TCATTATTAC TAACAGAGAT ACTTGTATCA TCATAAAGAC	5580
TCACTACTTG AGCATAAGAA GGCATCGTT GACTCAGATC CATTCTTGC CCATCTATAG	5640
TAATATTTGA CATGTTTATC CCAAAAGGAC TCTCCAAATA TTTAATAGCT TCTTCCCAA	5700
CTGTATCCGT GATATATAGT CAATTGAAAC AAGAGCAGGA TAAAAAGCC TCGTAAAGG	5760
TATTGCAACT TGGTAATACC TTTTGAGGT GCTTTTGTAT ATGAGCCCAT GTTTCTCAA	5820
TAGGATTGTA CTCAGGCGAG TAGGGAGGAA GAGGTAAAG TTTATGCCC AACTCTTCGC	5880
ATAAAAGTTC TAGCTTCCCC ATTCTATGGA ATCTTACATT ATCCATAATA ATAACCGATG	5940
GTGTGTTTAA TGTGGTAAG AGAAAATTCT GAAACCAAGC TTCAAAAAAG TCGCTCGTCA	6000
TCGTCTCTTC GTAAGTCATT GGAGCGATTA ATTCACCATT TGTTAGACCT GCAACCAAAG	6060

157

AAATCCTCTG ATATCTTCTT CCAGATACTT TGCCTCTTAT TAATTGACCT TTTAATGAGC	6120
GACCATATTC TCGATAAAAA TAAGTATCGA ATCCTGTTC GTCAATCTAA ACAGGTGCTA	6180
GGTGCTTTAA ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTCTGGG TCTGTTCAT	6240
AGTAGGTGTG GTTCTTTTTT CGAGTGTAGC CCATAGCTTT GAGCGTATAG TGGATGGTAG	6300
TTGGATGACA GCCAAATTCA GAAGCTATTT CAGTCAAATA AGCGTCTGGA TTGTCAGTAA	6360
GATAGTTTTT AAGTCTATCT CTATCAACCT TTCTTGGTTT TATTCCTTTT ACTTGGTGGT	6420
TTAGCTCTCC TGTCTCTCT TTTAGCTTTA ACCAGCCATA AATGGTATTA CGTGAGATTT	6480
GGAAAACGTG TGATGCTTCT GTTATACTAC CTGTTGCTC ACAATAAGAG AGAACTTTTT	6540
TACGAAAATC TATTGAATAT GCCATAAAAA GATTATACCA CATTGTGTAC TATTTTGGT	6600
TCATTTTACT ATATTTGAAG AGGCGTTTAA ACTATCTGAC ATAAAACCTCG TTCTAGAGGA	6660
AAGACATCCT TAAAAAGTT AGTTTATTTT ACAACTTAGA CATCAAGGTA GGTAAACCCC	6720
TTCATGGAAA AATCAAGACT CTTAGCACTA TGGGTAAAC TACCACTGGA GACGTAATCA	6780
ATCGCTAAAC CACGAAAACG GCTAATAGTG GTCATATCAA TATTTCCAGA ACATTCAATC	6840
CGAGAACGTC CTGCAATTAG GGTAATGGCC TGTTCATCT GTTCCAATGA CATATTATCC	6900
AACATGATAA TATCAGCACC CGCCGCCGCA GCTTCTTCGG CAGCAGCAAG GCTTCCACT	6960
TCCACCTCGA CCATTTTCAC AAAAGGGGCA TAGGCACGCG CTTGAGCAAT TGCCTTTTGA	7020
ACACTACCTA CTGCCGCAAT GTGATTGTCT TTTAGCAGGA TAGCATCTGA TAAATTAAAG	7080
CGATGATTAT AGCCACCGCC AACTCTCAGC GCATATTTCT CAAAAAGACG TAAATTAGGA	7140
GTAGTTTTTC GAGTATCAAA TACCTTAATG CAATCATCGC CTAAGGCTTC TACATAAGCA	7200
GCTGTCAATCG AAGCAATCCC TGATAAATGT TGTAATAAAT TCAAGGCAAC GCGTTCACAT	7260
GTTAAGAGAC TTCTACCGA GCCTATGATT TCTAAAACCA AATCGCCACT AGTCAAACGA	7320
TCCCCATCCT TAAATTGATG AGGATTCTGG AAGGTCACCT CGGCATCAAA TAGGGTAAAA	7380
ACCCCTTGAA AAACGGTTAG CCCCCTAAA ACACCAGCTT CCTTGGCAAA AAGCGACACC	7440
TTGGCTTGGC CATGATGATC AAAAATGGCA TTGGTACTGT AATCTTCGGA ATGAACATCT	7500
TCTCGCAAGG CTGCTTTCAA TGTATCATCT ATTTGAAAAG GGGTTAAATC AGTTGAAATG	7560
ATTGACATCA C	7571

(2) INFORMATION FOR SEQ ID NO: 3:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26385 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

158

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

TTTGCTAGTG GCTTAAATTC TTCAGGAAAA TCAGGCGTAT CTAAAAGTCG TGTCGTTTTT	60
GTTTCATCTA TATAAAGACT TCCTGCTCCC CCTACAACTA GAAAACGTGT CTGTGTTCCA	120
GCAAGAAGCT GATTAAATAG TTCGATTGAT TTGCTGTGGA GCGGTAGCGT ATCTGGTGTA	180
TAAGCACCAA ACGCTGAAAT AACAGCATCA AATCCAGTAA GATCATCTTT TGTCAACTCA	240
AATAATCTT TTTTAATAAT AGACTCAGCT TGACTTTTGT TTTCAGAACG AACATAGCC	300
GTTACTTCAT GTCCTCGTTT GACTGCTTCT TCAACAAATG CTTTCCCCGC TTGTCCATT	360
GCTGCAATAA CTGCTAGTTT CATTTTTTAT ACCTCTCTTG TTGTAATTAT TTTAGTTACA	420
GAAATTGTGA CACTCTTAAT AATCAATGTC AATAGTCTTG CTTAATTATT ATCAAAATAT	480
TTCTACCAAG AAAACTAACC ATGATTCTAG TGAAAAAAA TCTTCTTTGT CAACAAATTT	540
ACTTCTTGT TTTAAACATG CTATAATAAT CATAGCAAGA GATCTAAGTT GTCTGTTTTT	600
TTAAACGAG GTGATTATCA TCGGTAGATT CTATTCCCAT CTCCCCTACT ATCTGGTCAT	660
ATTATCTTT TATTGGCCAC TTTATGAGTT GTTCTTACTA GTTGTTCCTG ACCCCCTTAC	720
ACTCAAGGGA CTCTATATAA ACAATCTTCT CTTCTTTACA CCTCTGGTAA TCTTGATTGT	780
ATCGTTACTC TATAGCTACC GTTCCGTTT CTCACCTTGA TGGTTAGTTG GTAACGGACT	840
GCTCTTTTAC TTTACTATCA TAACCTTGG TGAGTTTATA CTAATTTACT TGCTAATCTA	900
TGAAACAGTT GCTCTGGTCG GCATGGATTC TGGTATTAGC ATCAAGCATA TTCTACAAAA	960
AATGAAAAAC AAAAACTTT CACAAAATCC TTGAAAAATC TCACAATCAT GCTATAATAA	1020
TCCATAGAGA CAAGTCACTT AGTCCCTTTC TACTAGAGAG TGCGTGGTTG CTGGAAACGC	1080
ATAGGAAGTC TAAACTGATA CTACTCTTGA GTTTTTTATG AAAACATAAA ACGGTGGCCA	1140
CGTTAGAGCC GATCAGAGGT GTCCCTCTCT TTTGAGGTAC ATAAATGAAG GTGGAACCA	1200
GTTGCGACGT CCTTTCGAGG ATGTCGCATT TTTTATTAG GATACTAATT ATGGAGTTGC	1260
AAGAATTAGT GGAGCGCAGT TGGGCAATCC GACAAGCTTA TCACGAACTG GAAGTTAAGC	1320
ATCATGATTC CAAGTGGACG GTAGAAGAAG ACCTCTTGGC TTTATCTAAT GATATTGGAA	1380
ATTTCCAACG ACTGGTGATG ACAAAGCAAG GACGCTACTA TGATGAAACA CCCTACACAC	1440
TGGAACAAAA ACTTTCAGAA AATATCTGGT GGCTATTAGA ACTTCTCAA CGTTTGGATA	1500
TAGACATTCT GACGGAATG GAAAACCTCC TCTCTGATAA AGAAAAGCAA TTGAACGTTA	1560
GGACTTGGAA GTAGTCTGCT GATAAAAAAT CAATGCTTAG AAACATATGAA ATAATAAAAA	1620

AGGAGAACAT CATGATTAAC ATTACTTTCC CAGATGGCGC TGTTTCGTGAA TTCCAATCTG	1680
GCGTAACAAC TTTTGAAATT GCCCAATCTA TCAGCAATTC CCTAGCTAAA AAAGCCTTGG	1740
CTGGTAAATT CAACGGCAAA CTCATCGACA CTACTCGCGC TATCACTGAA GATGGAAGCA	1800
TCGAAATTGT GACACCTGAT CACGAAGATG CCCTTCCAAT CTTGCGTCAC TCAGCAGCTC	1860
ACTTGTTCGC CCAAGCAGCT CGTCGTCTTT TCCCAGACAT TCACTTGGGA GTTGGTCCAG	1920
CCATCGAAGA TGGTTTCTAC TACGATACTG ACAACACAGC TGGTCAAATC TCTAACGAAG	1980
ACCTTCCTCG TATCGAAGAA GAAATGCAAA AAATCGTCAA AGAAAACTTC CCATCTATTC	2040
GTGAAGAAGT GACTAAAGAC GAGGCACGTG AAATCTTCAA AAATGACCCT TACAAGTTGG	2100
AATTGATTGA AGAACACTCA GAAGACGAAG GCGGTTTGAC TATCTATCGT CAGGGTGAAT	2160
ATGTAGACCT CTGCCGTGGA CCTCACGTTC CATCAACAGG TCGTATCCAA ATCTTCCACC	2220
TTCTCCATGT AGCTGGTGCG TACTGGCGTG GAAACAGCGA CAACGCTATG ATGCAACGTA	2280
TCTACGGTAC AGCTTGCTTT GACAAGAAAG ACTTGAAAAA CTACCTTCAA ATGCGTGAAG	2340
AAGCTAAGGA ACGTGACCAC CGTAACTTG GTAAAGAGCT TGACCTCTTT ATGATTTTAC	2400
AAGAAGTGGG ACAAGGTTTG CCATTCTGGT TGCCAAATGG TGGGACTATC CGTCGTGAAT	2460
TGGAACGCTA CATCGTAAAC AAAGAGTTGG TTTCTGGCTA CCAACACGTC TACACTCCAC	2520
CACTTGCTTC TGTTGAGCTT TACAAGACTT CTGGTCACTG GGATCATTAC CAAGAAGACA	2580
TGTTCCCAAC CATGGACATG GGTGACGGG AAGAATTTGT CCTTCGTCCA ATGAACGTGTC	2640
CGCACCACAT CCAAGTTTTC AAACACCATG TTCACTCTTA CCGTGAATTG CCAATCCGTA	2700
TCGCTGAAAT CCGTATGATG CACCGTTACG AAAAACTCGG TGCCCTCACT GGCCTTCAAC	2760
GTGTACGTGA AATGTCACTC AACGACGGTC ACCTATTCGT TACTCCAGAA CAAATCCAAG	2820
AAGAATTCCA ACGTGCCCTT CAGTTGATTA TCGATGTTTA TGAAGACTTC AACTTGACTG	2880
ACTACCGCTT CCGCCTCTCT CTTCTGTGACC CTCAAGATAC TCATAAGTAC TTTGATAACG	2940
ATGAGATGTG GGAAAAATGCC CAAACCATGC TTCGTGCAGC TCTTGATGAA ATGGGCGTGG	3000
ACTACTTTGA AGCCGAAGGT GAAGCAGCCT TCTACGGACC AAAATTGGAT ATCCAGATTA	3060
AAACTGCCCT TGGAAAAGAA GAAACCTTTT CTAATATCCA ACTTGATTTT TGTGTGCCAG	3120
AACGCTTCGA CCTCAAATAC ATCGGAGCTG ATGGCGAAGA TCACCGTCCA GTCATGATCC	3180
ACCGTGGGGT TATCTCAACT ATGGAACGCT TCACAGCTAT CTTGATTGAG AACTACAAGG	3240
GGGCCTTCCC AACATGGCTG GCACCACACC AAGTAACCTT CATCCCAGTA TCTAACGAAA	3300
AACACGTGGA CTACGCTTGG GAAGTGGCCA AGAAACTCCG TGACCGCGGT GTCCGTGCAG	3360

160

ACGTAGATGA GCGCAATGAA AAAATGCAGT TCAAGATCCG TGCTTCACAA ACCAGCAAGA	3420
TTCCTTACCA ATTAATTGTT GGAGACAAAG AAATGGAAGA CGAAACAGTC AACGTTTCGT	3480
GCTACGGCCA AAAAGAAACA CAACTGTCT CAGTTGATAA TTTTGTTCAA GCTATCCTAG	3540
CTGATATCGC CAACAAATCA CGCGTTGAGA AATAAGAGTC TAGCATAAAA GCCTCCAATC	3600
TGGAGGCTTT TTCTCATCTA TTTTACTCA AGGACTAAGT TCACTTGAGC AAAGTGAATC	3660
CGCACTGTCG TTCCTTTTCC GACCTCAGAC TCGATACGAA TCTGGTGCCC CAGTTCTTCA	3720
GAAATTTTCT TAGATAGATA AAGGCCAAGT CCAGAGGACT GCTGGGTCAA ACGGCCATTG	3780
TATCCTGAAA AGCCACGTTT AAATACTCGG AGGACATCAC TGTTTTTTAT CCCGATTCCC	3840
GTATCTTTGA TACAAAGCTC TTGGTCATCC ATATAAATCT CCAGACCACC TTCCTTGGTG	3900
TACTTGAGAC TGTTTGAGAT GATTTGCTCA ATAACCACTA GCAGCCACTT TTTATCCGTC	3960
ACGATTTCTT TATCAAGGTC ATGTAGATTG ACATTTAAGC CTTTTTGAAT AAAGAAAAGA	4020
GCAATTTTAC GAATTATTTT CTTGACCAAG TCCTCAATTT GAACCTGCTT TAAGACCAAA	4080
TCATCATGGA AACTTTCTAA ACGCAGGTAC TGTA AAACTA GGTGTTGATA GGAGTCGATT	4140
TTGAAAATTT CTGTCTTAG CTGCTGCTC AGTTGGCGGT CGACCACTTC TGCAACTAAG	4200
AGTTGACTGG CTGCAATGGG GGTCTTTATC TGATGGACCC ACAAGGTATA GTAATCCAGC	4260
AAATCCGTCA GTTTCTTTC TGCTTTTGAC CTCTGCTGAT AGAGTTCCAT CTCACGCGCT	4320
TCTAATTTT CTGCTAAAGC TATTTCCAAA GGAGACTTGG CTTCCCTCTC TCCATAGAGA	4380
AGTTCCTGGC GATAGACCTG CGTTTCCACC AATATGTCCC AAGTGAAAAA TAATATGTTT	4440
ACAAAGCAAC ACAAGAAGAA AAAGTAGAGG AAGTAAATTC CTAGACTGGC AAATAAAAC	4500
TGAAAGAGTA AGACAAGAAA TGCCAAAGAA AGCAGATAGA TAAAAAGACG ACTACGGGAG	4560
CGCAGATAGG CTAGAAAAAA TTGTTTCCAA TCAAGCATGC TTCAATCCGT ACCCTATTCC	4620
TTTCTTGGTC TCGATAAATC CTACCAATCC CTGCTCCTCC AACTTTTAC GCAAACGAGC	4680
CACATTGACA GAGAGGGTAT TATCATCAAT GAAAAAGTCA CTGTTCCAAA GTTCCCGCAT	4740
CAGGTCGTCA CGTGCTACGA TGTGCTGCTC ATGCTCAAAT AACACGCGTA AAATCTGGAA	4800
TTCAATCTTG GTCAAATTCA AGACTTGCCC TTGATAATGT AAATCCATGG ATTTGGTATT	4860
GAGGATAACA CCAGCATATT CCAGCAAACCT CTCATCACGC CCAAACCTCAT AGGAACGACG	4920
CAACAAGCCC TGAACCTTAG CTAAGAAGAAC CTGCTGGTCA AAAGGCTTGG TCACAAAGTC	4980
ATCCGCCCC ATATTGATTG CCATGACAAT ATCCATAGCC TGGTCTCTCG AAGAAAGAAA	5040
CATGATAGGT ACCTTGGAAT TCTTGCGGAT TTCCTGACAC CAGTGATAAC CATTAAACAA	5100
GGGCAAAACCA ATATCCATGA GGACCAGATG AGGTTCGGAC TGAACAAATA GACTCAAAAC	5160

161

TTCCATAAAG TCTTCTACCA GGACCACTTC AAATCCCCAT TCAGAGAGCA TTTTCCCAAT	5220
CTGTTGACGA ATGACCTGAT CATCTTCTAT TAATAAAATC TTGTGCATGC GCTTCTCCTT	5280
TTCCATTATT ATAACAGATT TTTCCATGCT AGATGGTCTG AAAGTGAATT TGAAATAGCC	5340
TGTTTTTAGC CAGTACAAAC AGGCTATGCT ACTAGCTAAT TTGAGGGAAA TTTGCTAAGA	5400
TAAATAAAAA GAAAGGAGCT CTTATGGCCA ATATTTTTGA CTATCTGAAA GATGTCCGAT	5460
ATGATTCTTA TTACGACCTT CCCTTGAATG AGTTAGACAT TCTAACCTTA ATAGAAATCA	5520
CCTACCTCTC CTTTGATAAT CTGGTCTCCA CACTTCTCA ACGTCTTTTA GATCTAGCAC	5580
CTCAGGTTCC AAGAGATCCC ACCATGCTTA CTAGCAAAAA TCGCCTTCAA TTATTAGATG	5640
AATTGGCTCA ACACAAGCGC TTCAAAAATT GCAAACTCTC CCATTTTATC AACGACATCG	5700
ACCCTGAACT GCAAAAGCAA TTTGCGGCTA TGACTTATCG TGTCAGCCTC GATACCTATC	5760
TGATTGTCTT TCGTGGGACA GATGACAGTA TCATTGGCTG GAAGGAAGAT TTCCACCTGA	5820
CCTATATGAA GGAAATTCCT GCTCAAAAGC ACGCCCTTCG CTATTTAAAG AACTTTTTTG	5880
CCCATCATCC TAAGCAAAAG GTTATTCTAG CTGGGCATTG CAAGGGAGGA AATCTCGCTA	5940
TCTATGCTGC TAGCCAAATT GAGCAAAAGT TGCAAAATCA GATCACAGCA GTTTATACAT	6000
TTGATGCACC TGGTCTCCAT CAAGAATTGA CACAGACTGC GGGTTATCAA AGGATAATGG	6060
ATAGAAGCAA GATATTCATT CCACAAGGTT CCATTATCGG TATGATGCTG GAAATTCCTG	6120
CTCACCAAAT CATCGTTCAG AGTACTGCCC TGGGTGGCAT CGCCAGCAC GATACCTTTA	6180
GTTGGCAGAT TGAGGACAAG CACTTCGTCC AACTGGATAA GACCAACAGT GATAGCCAGC	6240
AAGTAGACAC AACCTTTAAA GAATGGGTGG CCACAGTCCC TGACGAAGAA CTTGAGCTCT	6300
ACTTCGACCT CTTCTTTGGC ACTATTCTTG ATGCTGGTAT TAGCTCTATC AATGACTTGG	6360
CTTCCTTAAA GGCGCTTGAA TACATTCATC ATCTCTTTGT CCAAGCTCAA TCCCTCACTC	6420
CAGAAGAAAG AGAAACCTTG GGTGCGCTTA CCCAGTTATT GATTGATACT CGTTACCAGG	6480
CATGGAAAAA TAGATAATAC TCTTGAAAAT TAAATGTATA CAAAAACAAA GACCTAGAAT	6540
ACATACTTTC ATGTGCATTG TAAGTCTTTT TAAATAGAAT CTAATAGTCA ATAAAAATCA	6600
AAGAGCATTG AGAGATAATG GGGCTTGGAA CGTCCCTCTC GCTTCAACAA AATGACCCCA	6660
TTATAGATT A AAAGATGCC ACTTAGAAAA AGCAAAAAAG GAAGTAAGAC AAAGGCAAAT	6720
ATATAAAAAAG CTAAGTGAAC ATTCTCGTAT CCATTTTAT AAAAAAGGTA GGATAGATAA	6780
AAATAACTTG AAATGAGGGA TAATAAAAAA AATACTGGAT TCCACAACT TCTATTATCC	6840
TTCCAAAATG AACTATATAA GGCTAATACA ATTCCTATAA CGAGATACAT TTCTTACTCC	6900



162

TTTAATAGCT ACATTTTATC ATAATTATCC AAAGAAAAAA GAGGGCATT TATCCCTCTTA	6960
ATCCTTCATC TGA CTCTCTG CATCGGCCAC GACTTTTCT AGACTGGTTT GACCAAGTTC	7020
TGCCTCCATA GTCAACTGAA TTCTCTCAA TTTTGTATCC AAAACATCAT GAATATGAGC	7080
TCCTACAGGG CAATTTGGAT TCGGATTGTC ATGGAACTG AAGAGTTGAC CTGTCTTACC	7140
AAGACATTCG ACCGCCTGAT AAACATCTAA AAGACTAATA TCTTTAAGGT CCTTGACAAT	7200
CTCTGTTCGG CCCGTTCAC GCGCTACTGA AATCAGCTCT GCCTTCTTCA ACTGGGACAA	7260
GATCTTTCTG ATAATGACAG GATTGACCCC GACACTAGCA GCCAGAAAAT CACTGGTCAC	7320
CTTGCTTTCC TCCCCCTGA GGGCAATGAT TATCAGCATA TGAGTCGCAA TGGTAAATCT	7380
ACTTGAATT TGCATCCTCT TCTCCTTTT ACAGGGCTAC CCTGCCCTCA CTCTCTTTT	7440
TCTATTATTA TACCCTTTT AGTTGTAATG TCAATCGTTA CCACTTTTCA ACCAGTCGTC	7500
TAAC TCCCGA TCGCAGCCCT CTTCTGAGC CAATTCTCTC AAAAATTCCT GATGATGAGT	7560
ATGGTGGATC CCATTGACCA GACTTTCATA GTAAACCTCA AAATAGGGAA GTCTCAGGTC	7620
TTTAGCCAGC TGCAATTCAG CTGCTACATC GTAGTCTACC CGTCGGAGT CCATATCTAC	7680
CAGGCCTTTG TCATCAAAC CCAAAATCAT ATACTGGGCC CGCAAGTCCT TCCGTAGCTG	7740
AGCGTCCAAA AAGAAAGGTT GGCCAATCGA ACCCGGATTG ACAATCAATT GCCCACCAGT	7800
CCCCTAACGA AGCAACTGCT GGTGAATATG TCCATAAACA GCAATATCAC AGGGAGGATG	7860
AGTCACCAAG CGGTCAAAC CCTCTTGT TCCAGTATGA ATCAACTCTC GCCCCAGTT	7920
CTTATCAGGC AGATGATGGC TAATTCAC CGTCAAATCC CCAAACGAC GATGAATTTG	7980
AAGAGGTTGA TTGTGGAGCA CTTCAATTC TTCTAGGGAA ATTTCTCTA AAACATACTG	8040
GCACTGGCC AAGAGATAGC GTTGA CTGG GCGAGTACTG TCCAATTCCT TACGGACACC	9100
ATGCCAAAGA CTGTCTTCCC AGTTTCCCAA AACTCTAGCC GTAATCGGTA GTTGATCCAA	8160
CAAGTCCAAA ATCCTTCTAC GCCCTGTCCC TGGCATGAGA ATATCTCCCA AAAGCCAGTA	8220
TTTATCCACT CCTATCTGCC GAGCATCTGC CAAAACAGCC TCCAAGGCGG TGGTATTTCC	8280
ATGAATATCT GAAAGAAGAG CTATTTTCGT CATATCCATC TCCTCGTTT TTCTCTTGCA	3340
ATAAGTATAA CATAAAAAGT CACAGCTAGA GAAATCTAGC TTTTCTTGAT ATACTAGATA	9400
AAGATATTAG ACAAGAGGAA ACGAATGACC CAAAACAAAG AAGACTATCT AAAATGTATT	8460
TATGAAATTG GCATAGACCT GCATAAGATT ACCAACAAAG AAATTGCGGC TCGCATGCAA	8520
GTCTCTCCCC CTGCCGTAACT TGAAATGATC AAACGAATGA AAAGTGAAAA TCTCATCCTA	3580
AAGGACAAGG AATGTGGCTA TCTACTGACT GACCTCGGTC TCAAACGGT CTCTGAGCTC	8640
TATCGTAAGC ACCGCTTGAT TGAAGTTTT CTAGTTCATC ATTTAGACTA TACAAGTGAC	8700

163

CAGATTACAG AGGAAGCTGA GGTCTTGGAA CACACTGTCT CTGACCTGTT CGTGGAAGA	8760
CTAGATAAAC TGCTAGGTTT CCCTAAAACC TGCCCCCAGG GGGGAACTAT TCCTGCCAAG	8820
GGAGAACTAC TCCTTGAAAT CAATAACCTC CCACTAGCTG ATATCAAGGA AGCTGGCGCC	8880
TACCGCCTGA CTCGGGTGCA CGATAGTTTT GACATTCTCC ATTATCTGGA CAAGCACTCA	8940
CTTCACATCG GTGACCAGCT CCAAGTCAAG CAGTTTGATG GCTTCAGCAA TACCTTCACT	9000
ATCCTCAGTA ACGACGAGGA TTTACAAGTG AATATGGACA TTGCAAAACA ACTCTATGTC	9060
GAGAAAATCA ACTAATTCTT CAAGTCCCTT ACCAACCTTG AAAGTTTAT TTTGGCTCTT	9120
TGTCAACTGT AGTGGGTGA AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCTTTCTT	9180
TTTTTGATAT TCAGAGCGAT AAAATCCGT TTTTGAAGT TTTCAAAGT CCGAAAACCA	9240
AAGGCATTGC GCTTGATAAG TTTGATGAGA TTATTGGTCG CTTCCAGTTT GGCATTAGAA	9300
TAGTGTAGTT GAAGGGCGTT GACAATCTTT TCTTTATCTT TGAGGAAGGT TTTAAAGACA	9360
GTCTGAAAAA TAGGATGAAC CTGCTTTAGA TTGTCTCAA TGAGTCCGAA AAATTTCTCC	9420
GGTTTCTTAT TCTGAAAGTG AAACAGCAAG AGTTGATAGA GCTGATAGTG GTGTTTCAAG	9480
TCTTGTAAT AGCTCAAAAG CTGTCTAAA ATCTCTTAT TGGTTAAGTG CATACGAAAA	9540
GTAGGACGAT AAAATCGCTT ATCACTCAGT TTACGGCTAT CCTGTTGTAT GAGCTTCCAG	9600
TAGCGCTTGA TAGCCTTGTA TTCATGGGAT TTTGGATCCA ATTGGTTCAT AATTTGAACA	9660
CGCACACGAC TCATAGCAGG GCTAAGATGT TGTACAATGT GAAAGCGATC CAACACGATT	9720
TTAGCATTCG GGAGTGAAAC AGTCTGGGAG ACTGTTTCAG CCTGAGCCTA GAAATTTGAA	9780
AGCGAAGCTG TTTAGCCAAG TCATAGTAAG GACTAAACAT ATCCATCGTA ATGATTTTCA	9840
CTTGACAACG AACGGCTCTA TCGTAGCGAA GAAAGTGATT TCGGATGACA GCTTGTGTTT	9900
TGCCTTCAAG AACAGTGATA ATATTAAGAT TATCAAAATC TTGCGCAATG AAACCTCATCT	9960
TTCCCTTAGT GAAGGCATAC TCATCCCAAG ACATAATCTT TGGAAGCCGA GAAAAATCAT	10020
GCTCAAAAGT AAAGTCATTG AGCTTGGCAA TGACAGTTGA AGTTGAAATG GCCAGCTGAT	10080
GGGCAATATC AGTCATAGAA ATTTTTTCAA TTAACTTTTG AGCAATyTTT TGGTTGATGA	10140
TACGAGGGAT TTGGTGATTT TTCTTTACCA GGGGAGTCTC AGCAACCATC ATTTTTGAAC	10200
AGTGATAGCA CTTGAAACGA CGCTTTCTAA GGAGAATTCT AGAAGGCATA CCAGTCGTTT	10260
CAAGATAAGG AATTTTAGAA GGTMTTGAAG AGTCATATTT CTTCAATTGG TTTCCGCACT	10320
CAGGGCAAGA TGGGGCGTCG TAGTCCAGTT TGGCGATGAT TTCCTTGTGT GTATCCTTAT	10380
TGATGATGTC TAAAATCTGG ATATTAGGGT CTTTAATGTC TAGTAATTTT GTGATAAAAT	10440

164

GTAATTGTTT CATATGATTC TTTCTAATGA GTTGTTTTGT CGCTTTTCAT TATAGGTCAT	10500
ATGGGACTTT TTTTCTACAA TAAAATAGGC TCCATAATAT CTATAGTGGA TTTACCCACT	10560
ACAAATATTA TAGAACCGTA AAAATAGAAG GAGATAGCAG GTTTTCAAGC CTGCTATCTT	10620
TTTTTGATGA CATTTCAGGCT GATACGAAAT CATAAGAGGT CTGAAACTAC TTTTCAGAGTA	10680
GTCTGTTCTA TAAAATATAG TAGATTGAAA TAAGATGTGA ACAACTCTAT CAGGAAAGTC	10740
AAATTAATTT ATAGAATTAT TTTAGCAGTC AAGGTGTAAT GTTATAGATT CAATATATTA	10800
TATGACTATT AACCTTGTCT TCTCCTAAAA TTGACTTTCT TGTCTTCTTA TCTTGTCCAC	10860
TCGAAACAAG TATTGTAAGA ATTTGATTAT TTTTGAAAGT ACTTTTAATA TACTTGATAT	10920
AGTTAAAAAA GATTTGAAAC TAAATTCCTA ATTAGAAAAA GACTTCAAAT ACTAAAAAA	10980
AAAAAGTATA CTCTAATTGA AAACGGTAAC AAACTAATT TAGAGAATGA AATATAGAGT	11040
ATTTCTCTCT TAAAAGTTTT TGGTGAAACG AGATGTAGAA AGGAGATTTA GCCAAAGACT	11100
CTATTAGTGC TAGAATAATA GATTAGAATT ATTTTAGAAA AACGAAGTGA GCAGCTTATA	11160
AATTCAAGTC CCCAAATAGA TTCATACTAG TATCTTTTGC AAAAAATAAA GGGCGACTTC	11220
CTTCATGAAT ATCAATTTC TCTATAAGGA AGGTAGCTAA TTGAACTAAC TTATTTATTC	11280
TGTTTGTGCG TAGAAAAATC AGACCTCCTT GTGAAGATTG AGGAGATACT TAATGAAAAT	11340
CAAAGAAGAA ACTAGCAAGC TAGTAGCAGA TTGCCCCAAA CACCGCTTTG AGGTTGTAGA	11400
TAAGACTGAC CTATATAATC CAAGGTGAAG CGACTGTGGT TTGAAGAGAT TTTCAAAGAG	11460
TATAGGCTAG AGAGTAGTGT TTTTATGTCC TTCTAGTAGA AAATGCTAGA CAGAAGAATG	11520
GGGAACCTGG ATAGGAAAAA TAGATTGAGA AAGGAGGTTA GAAGAGATGA TTATTACAAA	11580
AATTAGCCGT TTAGGAACCT ATGTGGGAGT AAATCCACAT TTTGCAACAT TAATAGATTT	11640
TCTAGAAAAA ACAGGACTAG AAAATTTAAC AGAAGGTTCTG ATTGCTATCG ATGGTAATCG	11700
ATTGTTTGGG AATTGCTTTA CTTATCTAGC AGATGGTCAA GCAGGGGCTT TCTTTGAAAC	11760
CCACCAAAAA TATTTGGATA TTCATTTAGT TTTGGAAAAA GAAGAAGCCA TGGCTGTTAC	11820
ATCGCCGAA AATGTAAGCG TTACCCAAGA ATATGATGAA GAGAAAGATA TTGAATTATA	11880
CACAGGGAAA GTGGAACAGT TGGTTCATTT GAGAGCTGGC GAATGCCTCA TCACTTTTCC	11940
AGAAGATTTA CATCAACCCA AGGTTCTGAT AAATGATGAA CCTGTGAAAA AAGTTGTCTT	12000
TAAAGTTGCG ATTTCTTAAT GTAGAAAGAG AAGAACGATG AAAAAAATGA GAAAGTTTTT	12060
ATGTCTAGCT GGAATTGCGC TAGCGGCTGT TGCCTTGGA GCTTGTTCAG GAAAAAAGA	12120
AGCTACAAC AGTACTGAAC CACCAACAGA ATTATCTGGT GAGATTACAA TGTGGCACTC	12180
CTTTACTCAA GGACCCCGTT TAGAAAGTAT TCAAAAATCA GCAGATGCTT TCATGCAAAA	12240

165

GCATCCAAAA	ACGAAAATCA	AGATTGAAAC	ATTTCTTTGG	AATGACTTCT	ATACTAAATG	12300
GACTACAGGT	TTAGCAAATG	GAAATGTGCC	AGATATCAGT	ACAGCTCTTC	CTAACCAAGT	12360
AATGGAAATG	GTCAACTCAG	ATGCTTTGGT	TCCGCTAAAT	GATTCTATCA	AGCGTATTGG	12420
ACAAGATAAA	TTTAACGAAA	CTGCCTTAAA	TGAAGCAAAA	ATCGGAGATG	ATTACTACTC	12480
TGTTCTCTTT	TATTCACATG	CACAAGTCAT	GTGGGTTAGA	ACAGATTTGT	TAAAAGAACA	12540
TAATATTGAG	GTTCTTAAAA	CTTGGGATCA	ACTCTATGAA	GCTTCTAAAA	AATTGAAAGA	12600
AGCTGGAGTT	TATGGCTTGT	CTGTTCCGTT	TGGAACAAAT	GACTTAATGG	CAACACGTTT	12660
CTTGAAC TTC	TACGTACGTA	CTGGTGGAGG	AAGCCTCTTA	ACAAAAGATC	TTAAAGCAGA	12720
CTTGACAAGC	CAACTTGCTC	AAGATGGTAT	TAAATACTGG	GTAAATTTGT	ATAAAGAAAT	12780
CTCACCTCAA	GATTCTTTGA	ACTTTAATGT	CCTTCAACAA	GCTACCTTGT	TCTATCAAGG	12840
AAAAACAGCA	TTTGACTTTA	ACTCTGGCTT	CCATATCGGA	GGAATTAATG	CCAACAGTCC	12900
TCAATTGATT	GATTGATTG	ATGCTTATCC	TATTCAAAAA	ATCAAAGAGT	CTGATAAAGA	12960
CCAAGGAATT	GAAACCTCAA	ACATTCCAAT	GGTTGTTTGG	AAAAATTCAA	AACATCCAGA	13020
AGTTGCTAAA	GCATTCTTAG	AAGCACTTTA	TAATGAAGAA	GACTACGTTA	AATTCCTTGA	13080
TTCAACTCCA	GTAGGTATGT	TGCCAACTAT	TAAGGGGATT	AGCGATTCTG	CAGCCTATAA	13140
AGAAAATGAA	ACTCGTAAGA	AATTTAAACA	TGCTGAAGAA	GTAATTACTG	AAGCTGTTAA	13200
AAAAGGTA CT	GCTATTGGTT	ATGAAAATGG	GCCAAGTGTA	CAAGCTGGTA	TGTTGACTAA	13260
CCAACACATT	ATTGAACAAA	TGTTCCAAGA	TATCATTACA	AATGGAACAG	ATCCTATGAA	13320
AGCAGCAAAA	GAAGCAGAAA	AACAATTAAA	TGATTTATTT	GAGGCTGTTT	AGTAGATGTA	13380
AAAGACTAGA	AAATAGGTGG	GATAGTGAGC	TGAAAAGCTC	TAGCCCAATC	TTGTAAAAGA	13440
AGGGAGAAGG	AGAATGGTTA	AAGAACGTAA	TTTAACTCGC	TGGATATTTG	TTTTGCCAGC	13500
TATGATTATC	GTAGGATTAC	TCTTTGTTTA	TCCGTTTTTC	TCGAGTATTT	TTTATAGCTT	13560
TACCAATAAG	CATTTGATTA	TGCCTAATTA	TAAATTTGTT	GGTTTGGCTA	ACTATAAAGC	13620
TGTGCTATCA	GATCCCAACT	TCTTTAATGC	GTTCTTTAAT	TCAATTAAGT	GGACCGTTTT	13680
CTCATTAGTT	GGTCAAGTTT	TAGTAGGGTT	TGTATTGGCT	TTAGCTCTTC	ACAGAGTACG	13740
CCACTTCAAG	AAATTATATA	GGACATTATT	GATTGTTTCT	TGGGCATTTT	CTACCATCGT	13800
TATTGCCTTC	TCTTGGCAGT	GGATTCTAAA	CGGGGTTTAT	GGCTACTTAC	CTAATCTAAT	13860
CGTAAATTA	CGTTTAATGG	AACATACACC	TGCATTTTGT	ACAGATAGTA	CATGGGCATT	13920
CCTATGTTTG	GTGTTTATCA	ACATTTGGTT	TGGAGCACCA	ATGATTATGG	TTAATGTGCT	13980

166

TTCAGCTTTG CAAACAGTAC CAGAAGAACA ATTTGAGGCT GCTAAGATAG ATGGTGCTTC	14040
AAGTTGGCAG GTGTTCAAGT TTATCGTCTT TCCACATATT AAAGTGGTTG TAGGACTTCT	14100
AGTTGTTTTG AGAACTGTAT GGATCTTTAA TAACTTTGAC ATTATCTACC TCATTACTGG	14160
TGGTGGACCA GCCAATGCTA CAACGACGCT TCCAATTTT GCTTACAACC TGGGCTGGGG	14220
AACTAAATTG TTGGGTCGTG CTTCAGCAGT TACAGTACTG CTCTTTATCT TCTTGGTGGC	14280
GATTTGCTTT ATCTACTTTG CTATCATCAG TAAGTGGGAA AAGGAGGGTA GAAAATAATG	14340
AAGAAGAAAT CCAGTATTTA TTTAGATATT CTCTCACATG TACTTTTAGT TGGTGCGACC	14400
ATCGTTGCAG TTTTCCCAT TGGTATGGATT ATCATATCTT CTGTCAAAGG GAAAGGGGAA	14460
TTAACTCAGT ATCCAACACG ATTTTGGCCT GAACAGTTTA CATTAGATTA TTTCACATCAT	14520
GTTATCAACG ATTTGCACTT CATTGATAAC ATTCCGAAACA GTTTAATCAT TGCCTTGGCT	14580
ACAACCCTTA TTGCGATTAT TATTTCTGCT ATGGCAGCCT ATGGTATTGT TCGATTCTTT	14640
CCTAAATTGG GAGCAATCAT GTCGAGACTA CTCGTCATTA CCTACATTTT CCCACCAATT	14700
TTGTTAGCAA TTCCCTATTC AATTGCCATT GCTAAAGTTG GGTTAACAAA TAGTTTATTT	14760
GGCTTGATGA TGGTTTATCT ATCTTTTAGT GTTCCATATG CAGTTTGGCT CTTAGTTGGA	14820
TTTTTCCAAA CAGTTCCAAT TGGAAATGAA GAAGCGGCTA GAATTGATGG TGCAAAATAA	14880
TTTGTTACGT TTTATAAAGT TGTGCTACCG ATTGTAGCAC CAGGTATTGT AGCAACAGCT	14940
ATTTATACAT TTATCAATGC TTGGAATGAA TTCCTGTATG CCTTGATTTT GATTAACAAT	15000
ACAGGAAAGA TGACAGTAGC AGTAGCCCTT CGTTCACTTA ATGGTTCAGA AATACTAGAC	15060
TGGGGAGATA TGATGGCAGC GTCTGTTATT GTAGTTCTTC CATCAATTAT TTTCTTCTCT	15120
ATCATCCAAA ATAAGATTGC AAGTGGATTA TCAGAAGGAT CTGTGAAGTA GACGAAAGAA	15180
GGAAAAAAT GAATAAAAGA GGTCTTTATT CAAAAGTAGG AATTTCCGTT GTAGGCATTA	15240
GTCTTTTAAT GGGAGTCCCC ACTTTGATTC ATGCGAATGA ATTAACTAT GGTCAACTGT	15300
CCATATCTCC TATTTTTCAA GGAGGTTTAT ATCAACTGAA CAATAAGAGT ATAGATATCA	15360
GCTCTTTGTT ATTAGATAAA TTGTCTGGAG AGAGTCAGAC AGTAGTAATG AAATTTAAAG	15420
CAGATAAACC AAACCTCTCT CAAGCTTTGT TTGGCCTATC TAATAGTAAA GCAGGCTTTA	15480
AAAATAATTA CTTTTCATTT TTCATGAGAG ATTCTGGTGA GATAGGTGTA GAAATAAGAG	15540
ACGCCCAAAA GGAATAAAT TATTTATTTT CCAGACCAGC TTCATTATGG GGAACACATA	15600
AAGGACAGGC AGTTGAAAA ACACCTAGTAT TTGTATCTGA TTCTAAAGAT AAAACATACA	15660
CAATGTATGT TAATGGAATA GAAGTGTCTT CTGAAACAGT TGATACATTT TTGCCAATTT	15720
CAAATATAAA TGGTATAGAT AAGGCAACAC TAGGAGCTGT TAATCGTGAA GGTAAAGAAC	15780

167

ATTACCTCGC AAAAGGAAGT ATTGATGAAA TCAGTCTATT TAACAAAGCA ATTAGTGATC	15840
AGGAAGTTTC AACTATTCCC TTGTCAAATC CATTTCAGTT AATTTTCCAA TCAGGAGATT	15900
CTACTCAAGC TAACTATTTT AGAATACCGA CACTATATAC ATTAAGTAGT GGAAGAGTTC	15960
TATCAAGTAT TGATGCACGT TATGGTGGGA CTCATGATTC TAAAAGTAAG ATTAATATTG	16020
CCACTTCTTA TAGTGATGAT AATGGGAAAA CGTGGAGTGA GCCAATTTTT GCTATGAAGT	16080
<del>TTAATGACTA TGAGGAGCAG TTAGTTTACT</del> GGGCAGAGA <del>TAATAAATTA AAGAATAGTC</del>	<del>16140</del>
AAATTAGTGG AAGTGCTTCA TTCATAGATT CATCCATTGT TGAAGATAAA AAATCTGGGA	16200
AAACGATATT ACTAGCTGAT GTTATGCCTG CGGGTATTGG AAATAATAAT GCAAAATAAG	16260
CCGACTCAGG TTTTAAAGAA ATAAATGGTC ATTATTATTT AAAACTAAAG AAGAATGGAG	16320
ATAACGATTT CCGTTATACA GTTAGAGAAA ATGGTGTCTG TTATAATGAA ACAACTAATA	16380
AACCTACAAA TTATACTATA AATGATAAGT ATGAACTTTT GGAGGGAGGA AAGCTTTTAA	16440
CAGTCGAACA ATATTCCGTT GATTTTGATA GTGGCTCTTT AAGAGAAAGG CATAATGGAA	16500
AACAGGTTCC TATGAATGTT TTCTACAAAG ATTCGTTATT TAAAGTGACT CCTACTAATT	16560
ATATAGCAAT GACAACTAGT CAGAATAGAG GAGAGAGTTG GGAACAATTT AAGTTGTTGC	16620
CTCCGTTCTT AGGAGAAAAA CATAATGGAA CTTACTTATG TCCCGACAA GGTTTAGCAT	16680
TAAAATCAAG TAACAGATTG ATTTTGTCAA CATATACTAG TGGAGAACTA ACCTACTCTA	16740
TTTCTGATGA TAGTGGTCAA ACATGGAAGA AATCCTCAGC TTCAATCCG TTTAAAAATG	16800
CAACAGCAGA AGCACAAATG GTTGAAGTGA GAGATGGTGT GATTAGAACA TTCTTTAGAA	16860
CCACTACAGG TAAGATAGCT TATATGACTA GTAGAGATTC TGGAGAAACA TGGTCGAAAG	16920
TTTCGTATAT TGATGGAATC CAACAACTT CATATGGCAC ACAAGTATCT GCAATTAAAT	16980
ACTCTCAATT AATTGATGGA AAAGAAGCAG TCATTTTGAG TACACCAAAT TCTAGAAGTG	17040
GCCGCAAGGG AGGCCAATTA GTTGTGGTGT TAGTCAATAA AGAAGATGAT AGTATTGATT	17100
GGAAATACCA CTATGATATT GATTTCCTT CGTATGGTTA TGCCTATTCT GCGATTACAG	17160
AATTGCCAAA TCATCACATA GGTGTACTGT TTGAAAAATA TGATTCGTGG TCGAGAAATG	17220
AATTGCATTT AAGCAATGTA GTTCAGTATA TAGATTGGA AATTAATGAT TTAACAAAAT	17280
AAAGGAGAAA AACATGGTTA AATACGGTGT TGTGGAACA GGGTATTTTG GAGCTGAATT	17340
GGCTCGCTAC ATGCAAAAGA ATGATGGAGC AGAGATTACT CTTCTCTATG ATCCAGATAA	17400
TGCAGAGGCG ATTGCAGAAG AATTGGGAGC AAAAGTAGCA AGTTCCTTAG ATGAGTTGGT	17460
TTCTAGCGAT GAAGTAGATT GTGTTATCGT CGCAACTCCA AATAATCTTC ATAAGGAACC	17520

168

GGTTATTAAG GCTGCACAGC ATGGTAAAAA TGTTTTCTGT <del>GAAAA</del> CCAA TTGCGCTTTC	17580
TTATCAAGAT TGTCGCGAGA TGGTAGATGC GTGTAAAGAA AACAAATGTAA CCTTTATGGC	17640
AGGACATATT ATGAATTCTT TTAATCGTGT TCATCATGCA AAAGAACTCA TTAATCAAGG	17700
AGTTATCGGA GACGTTCTAT ATTGTCATAC AGCTCGTAAT GGTGGGAAG AACAAACAACC	17760
GTCAGTATCA TGGAAAAAAA TTCGTGAAAA ATCAGGTGGT CACTTGTATC ACCACATCCA	17820
TGAATTGGAT TGCCTTCAAT TCCTTATGGG GGGCATGCCT GAAACTGTAA CCATGACAGG	17880
TGGAAATGTG GCCCATGAAG GTGAACATTT CCGTGATGAA GATGATATGA TTTTGTCAA	17940
TATGGAATTT TCTAATAAGC GTTTTGCCTT GTTAGAATGG GGTTCAGCTT ATCGTTGGGG	18000
TGAACATTAT GTCTTAATCC AAGGAAGCAA AGGTGCCATC CGCTTAGACT TATTCAACTG	18060
TAAAGGAACT CTTAAGCTAG ATGGGCAAGA AAGCTATTTC TTGATTCACG AATCGCAAGA	18120
AGAAGATGAT GATCGGACTC GTATCTATCA TAGTACAGAG ATGGATGGAG CAATTGCTTA	18180
TGGTAAACCA GGTAAACGTA CTCCATTATG GCTATCATCT GTCATTGATA AAGAAATGCG	18240
CTATCTGCAT GAGATTATCG AAGGAGCTCC AGTATCAGAA GAATTTGCAA AACTTTTGAC	18300
AGGTGAAGCT GCCCTAGAAG CAATTGCTAC TGCAGATGCT TGTACCCAGT CTATGTTTGA	18360
AGATCGCAAA GTAAATTTGT CAGAAATTTGT AAAATAAATT TTGGTATTCT CCTATTTATA	18420
GGTCGACTTG CTCCTCTGAA AGTACTTTTA GAGGAGCTGT TTGACTTTGC TAGTTTTTGA	18480
AACTGAAATC TATTATACTA CAAACTATTG AAAGCGTTT AATTTTAAGG TATAATAATC	18540
TCATAGAAAT AAAGAAAAGG AGGAAAGAGG ATGCCACAGA TTAGCAAAGA AGCCTTGATT	18600
GAGCAAATCA AAGATGGAAT CATCGTTTCT TGTGAGGCTC TTCCTCATGA ACCGCTTTAT	18660
ACAGAAGCGG GAGGGGTGAT TCCCTTGCTG GTCAAAGCGG CTGAGCAAGG TGGAGCAGTC	18720
GGTATCCGAG CAAACAGTGT TCCGATATC AAGGAAATTA AGGAAGTCAC TAAACTTCCA	18780
ATCATTGGGA TTATCAAACG TGATTATCCA CCTCAGGAAC CCTTCATCAC GGCTACTATG	18840
AAAGAAGTTG ATGAATTGGC AGAACTGGAC ATCGAGGTGA TTGCTCTGGA TTGTACCAAG	18900
CGTGAACGCT ACGATGGTTT GGAAATTCAA GAGTTCATTG GTCAGGTAA CGAGAAATAT	18960
CCTAATCAGC TTTTGATGGC TGATACTAGT ATCTTCGAAG AAGGGCTAGC AGCTGTAGAA	19020
GCAGGAATTG ACTTTGTCGG AACAACTTA TCAGGCTACA CATCCTACAG TCCAAAAGTA	19080
GACGGTCCAG ATTTTGAATT GATTAAGAAA CTCTGTGATG CTGGTGTAGA TGTCATTGCA	19140
GAAGGAAAAA TTCATACACC AGAACAAGCC AAACAAATCC TTGAATATGG AGTGCGAGGC	19200
ATCGTTGTTG GTGGCCCAT TACTAGACCA AAAGAGATTA CAGAACGCTT CGTTGCTAGT	19260
CTTAAATAAG ATGTGAGGGG GAGTTTTATG TTTAAAGTTT TACAAAAAGT TGGAAAAGCT	19320

169

TTTATGTTAC CTATAGCTAT ACTTCCTGCA GCAGGTCTAC TTTTGGGGAT TGGTGGTGCA	19380
CTTTCAAACC CAACCACGAT AGCAACTTAT CCAATACTAG ACAATAGTAT TTTTCAATCA	19440
ATATTCCAAG TAATGAGCTC TGCAGGAGAG GTTGTATTCA GTAATTTGTC ACTACTTCTC	19500
TGTGTGGGAT TATGTATTGG CTTAGCGAAA CGAGATAAAG GAACCGCTGC GTTAGCAGGA	19560
GTAAGTGGTT ACTTAGTTAT GACTGCAACG ATCAAAGCTT TGGTAAACT TTTTATGGCA	19620
GAAGGATCTG CAATTGATAC TGGAGTTATT GGAGCATTAG TTGTCGGAAT AGTTGCCGTA	19680
TATTTGCACA ACCGATATAA CAATATTCAA TTACCTTCCG CTTTAGGATT CTTTGGAGGT	19740
TCACGCTTCG TTCCTATTGT TACATCGTTC TCTTCTATCT TGATTGGCTT TGTCTTCTTT	19800
GTTATTTGGC CACCTTTCCA ACAACTTCTT GTTCTACAG GTGGATATAT TTCTCAGGCG	19860
GGTCCAATTG GAACTTTTCT ATATGGATTT TTAATGAGAC TTTCTGGAGC AGTAGGCTTA	19920
CATCATATAA TTTACCCTAT GTTTTGGTAT ACTGAACTTG GTGGTGTGA AACTGTTGCA	19980
GGACAAACAG TGGTTGGAGC TCAAAAAATA TTTTTGCTC AATTAGCCGA TTTGGCCCAT	20040
TCTGGATTAT TTACAGAAGG AACCAAGGTTT TTTGCAGGTC GTTCTCAAC AATGATGTTT	20100
GGTTTACCGG CTGCCGTGTT AGCGATGTAC CATAGTGTTC CTAATAATCG TCGTAAAAAA	20160
TACGCGGGTT TGTTTTTTGG AGTTGCTTTA ACATCTTTTA TTACCGGTAT TACAGAACCA	20220
ATTGAATTTA TGTTTCTATT CGTCAGTCCG GTTCTATATG TTGTTACCGC ATTCCTTGAT	20280
GGTGTAGCT TCTTTATTGC AGACGTCTTA AATATTTCAA TAGGAAACAC ATTTTCAGGA	20340
GGTGTAAATG ATTTCACTTT ATTTGGAATT TTGCAGGGGA ACGCTAAGAC GAATTGGGTT	20400
CTTCAGATTC CATTGGACT TATTTGGAGT GTTTTGTATT ATATTATTTT TAGATGGTTC	20460
ATTACTCAAT TCAACGTTCT AACGCCAGGG CGAGGAGAAG AAGTAGATTC TAAAGAAATT	20520
TCTGAATCCG CAGATTCAAC TTCAAATACT GCAGATTATT TAAAACAGGA TAGCCTACAA	20580
ATTATCAGAG CCTTGGGTGG ATCAAATAAT ATAGAAGATG TAGATGCTTG TGTGACACGT	20640
TTACGTGTAG CTGTAAAAGA AGTTAATCAA GTTGATAAAG CACTTTTAAA ACAAAATTGGT	20700
GCAGTTGATG TCTTAGAAGT GAACGGTGGC ATTCAAGCAA TCTATGGAGC AAAAGCAATC	20760
TTATATAAAA ATAGTATTAA TGAAATTTTA GGTGTAGATG ATTAAGTACT TACTGACTTA	20820
ATAAAAAACA GAGGAGAGTG ATGGATGAGT AGGATGAAAT GAAATCGCAT ACAAGAAATA	20880
AAGAACTCAT TATCCAAGTT GGATACGCTT ATTACATAGG AGAATACAAA TGAAATTTAG	20940
AAAATTAGCT TGTACAGTAC TTGCGGTGTC TGCGGTCTT GGTCTGCTG CTTGTGGCAA	21000
TTCTGGCGGA AGTAAAGATG CTGCCAAATC AGGTGGTGAC GGTGCCAAAA CAGAAATCAC	21060



170

TTGGTGGGCA TTCCCAGTAT TTACCCAAGA AAAAAGTGGT GACCGTGTG GAACTTATGA	21120
AAAATCAATC ATCGAAGCGT TTGAAAAAGC AAEECCAGAT ATAAAAGTGA AATTGGAAAC	21180
CATCGACTTC AAGTCAGGTC CTGAAAAAAT CACAACAGCC ATCGAAGCAG GAACAGCTCC	21240
AGACGTACTC TTTGATGCAC CAGGACGTAT CATCCAATAC GGTAAAAACG GTAAATTTGGC	21300
TGAGTTGAAT GACCTCTTCA CAGATGAATT TGTAAAGAT GTCAACAATG AAAACATCGT	21360
ACAAGCAAGT AAAGCTGGAG ACAAGGCTTA TATGTATCCG ATTAGTTCTG CCCCATTCTA	21420
CATGGCAATG AACAAGAAAA TGTAGAAGA TGCTGGAGTA GCAAACCTTG TAAAAGAAGG	21480
TTGGACAAC TATGATTTTG AAAAAGTATT GAAAGCACTT AAAGACAAGG GTTACACACC	21540
AGGTTTCAATG TTCAGTTCTG GTCAAGGGGG AGACCAAGGA ACACGTGCCT TTATCTCTAA	21600
CCTTTATAGC GGTCTGTAA CAGATGAAAA AGTTAGCAA TATACAAC TGATCCTAA	21660
ATTCGTCAA GGTCTTGA AAGCAACTAG CTGGATTAAA GACAATTTGA TCAATAATGG	21720
TTCACAATT GACGGTGGG CAGATATCCA AAAGTTGCC AACGGTCAA CATCTTACAC	21780
AATCCTTTGG GCACCAGTC AAAATGGTAT CCAAGCTAAA CTTTATAAG CAAGTAAGGT	21840
AGAAGTGTA GAAGTACCAT TCCCATCAGA CGAAGGTAAG CCAGCTCTG AGTACCTGT	21900
AAACGGGTT GCAGTATCA ACAATAAAGA CGACAAGAAA GTCGCTGCAT CTAAGAAAT	21960
CATCCAGTT ATCGCAGATG ACAAGGAGTG GGGACCTAAA GACGTAGTTC GTACAGGTGC	22020
TTTCCCAGTC CGTACTTCAT TTGGAAAAC TTATGAAGAC AAACGCATGG AAACAATCAG	22080
CGGCTGGACT CAATACTACT CACCATACTA CAACACTATT GATGGATTG CTGAAATGAG	22140
AACACTTTGG TTCCCAATGT TGCAATCTGT ATCAAATGGT GACGAAAAAC CAGCAGATGC	22200
TTTGAAAGCC TCACTGAAA AAGCGAACGA AACAATCAA AAAGCTATGA AACAATAGTC	22260
CTTAGTTATT CTATAAAAAG TAGTTTTTTA AAGAACCCTAA GAGTGTATAC CCCCTTTTCC	22320
CTCTACACAG ATAGTGTAAG AAAAGGGGGC TTTTGTATAA AATGTAAGAA ACTGTACAG	22380
AATTAAAATG AAGTTCTTAC ATAAGCGAAT CATAAAAAAT TTCATTTTGA TTTTAAAAA	22440
GTTCAAGAAA GTCAAAAAAT TATCTATTT GAAAGAGAGG TGCCGACTGT GAAAGTCAAT	22500
AAAATCCGTA TGCGGGAAAC AGTGATTTCC TACGCTTTCC TAGCACCAGT ATTATTCTTC	22560
TTTGTCATCT TTGTGTTGGC TCCGATGGTG ATGGGCTTCA TTACAAGTTT CTTTAACTAC	22620
TCAATGACTA AATTTGAGTT TGTAGGCTTG GATAACTATA TCCGTATGTT TAAAGATCCT	22680
GTCTTTACAA AATCTCTGAT TAACACAGTT ATTTTGGTTA TTGGATCTGT ACCAGTTGTT	22740
GTCTATTCT CACTCTTTGT AGCATCTCAG ACCTATCATC AAAATGTCAT TGCCAGATCC	22800
TTCTACCGTT TCGTCTTCTT CCTTCCTGTT GTAACGGGTA GTGTTGCCGT GACAGTTGTT	22860

171

TGGAAATGGA TTTATGACCC ACTATCAGGG ATTCTAAACT TTGTCCTTAA GTCCAGCCAC	22920
ATCATCAGCC AAAACATTTC TTGGTTGGGA GATAAAAAC TGGCATTGAT GGCGATTATG	22980
ATTATTCTCT TGACCACTTC AGTTGGTCAG CCCATCATCC TTTATATCGC TGCCATGGGG	23040
AATATTGACA ATTCACTGGT TGAAGCGGCG CGTGTGATG GTGCAACTGA GTTCAAGTT	23100
TTTGGGAAGA TTAAATGGCC AAGCCTTCTT CCAACAAC TC TTTATATTGC AATCATCACA	23160
ACAATTAAC TATTCCAGTG TTTCCCTTG ATTCCAGCTTT TGACATCTGG TGGTCCAAAC	23220
TACTCAACAA GTACCTTGAT GTACTACCTT TACGAAAAAG CCTTCCAATT GACAGAATAC	23280
GGCTATGCCA ACACAATTGG TGTCTTCTTG GCAGTCATGA TTGCTATCGT AAGCTTTGTT	23340
CAATTTAAG TACTTGGAAA CGACGTAGAA TACTAAAGAA AGGAGACAGC TATGCAATCT	23400
ACAGAAAAA AACCATTAAAC AGCCTTTACT GTTATTTCAA CAATCATTTT GCTCTTGTG	23460
ACTGTGCTGT TCATCTTTCC ATTCTACTGG ATTTTGACAG GGGCATTCAA ATCACAACCT	23520
GATACAATTG TTATTCCTCC TCAGTGGTTC CCTAAAATGC CAACCATGGA AACTTTCCAA	23580
CAACTCATGG TGCAGAACCC TGCCTTGCAA TGGATGTGGA ACTCAGTATT TATCTCATTG	23640
GTAACCATGT TCTTAGTTG TGCAACCTCA TCTCTAGCAG GTTATGTATT GGCTAAAAA	23700
CGTTTCTATG GTCAACGCAT TCTATTTGCT ATCTTTATCG CTGCTATGGC GCTTCCAAAA	23760
CAAGTTGTCC TTGTACCATT GGTACGTATC GTCAACTTCA TGGGAATCCA TGATACTCTC	23820
TGGGCAGTTA TCTTGCCTTT GATTGGATGG CCATTCGGTG TCTTCCTCAT GAAACAGTTC	23880
AGTGAAAATA TCCCTACAGA GTTGCTTGAA TCAGCTAAAA TCGACGGTTG TGGTGAGATT	23940
CGTACCCTCT GGAGGTAGC CTTCCTGATT GTGAAACCAG GGTTCGAGC CCTTGCAATC	24000
TTTACCTTCA TCAATACTTG GAATGACTAC TTCATGCAAT TGGTAATGTT GACTTCACGT	24060
AACAAATTGA CCATCTCACT TGGGGTTGCG ACCATGCAGG CTGAAATGGC AACCAACTAT	24120
GGTTTGATTA TGGCAGGAGC TGCCCTTGCT GCTGTTCCAA TCGTCACAGT CTTCTAGTC	24180
TTCCAAAAAT CCTTCACACA GGGTATTACT ATGGGAGCGG TCAAAGGATA ATACTCTGCG	24240
AAAATCTCTT CAAACTACGT CAGCTTCACC TTGCCATACT TAAGTATTGC CTGCGGTTAG	24300
CTTCCTAGTT TGTCTTCAA TTTTCATTGA GTATAGGAAA ATCAATCTAT CAAGATACAG	24360
AAGTATATTT TATAGATTTA GAGAATATAG AGGTTATAAG TGTCTACAAA ATGGAGGGTA	24420
TGCAGTTACT TTATGAAGTT TTGTCAGACA CTTATAAACT TAAGAATGGT TTTAGTTAAC	24480
TATCAGAAAC GAAGGAAAGA GTATGATTTT TGACGATTTG AAAAACATCA CCTTTTACAA	24540
AGGGATTTCAT CCTAATTTAG ACAAGGCTAT CGACTATCTC TACCAACATC GTAAGGATTC	24600

172

TTTCGAATTA GGAAAGTATG ATATTGATGG AGATAAAGTC TTTCTAGTTG TTCAGGAAAA	24660
TGTCCTCAAT CAAGCTGAAA ATGATCAATT TGAGTATCAT AAGAACTATG CAGATTGCA	24720
TTTGCTGGTA GAAGGACATG AATATTCGAG CTACGGTTCA CGTATCAAAG ACGAGGCAGT	24780
AGCATTGAC GAAGCGAGTG ACATTGGCTT TGTTCATTGT CATGAACACT ACCCACTCTT	24840
GTTGGGTAT CACAATTTTG CGATTTTCTT CCCAGGTGAG CCACATCAGC CAAATGGTTA	24900
TGCAGGCATG GAAGAAAAGG TTCGAAAATA TCTCTTTAAA ATTTTGATTG ATTAAAAATA	24960
GGATGAATTG TTTTTTGT AAGCTTTGAT AATACTCTAC CATGAAATTG ATCTTTGTGA	25020
GGTAGAGAAA TGAGAATAAA ATATTTAAAA ATTGGTATCT TCTAAGTATG CTGCAAGAGC	25080
TAGTTTCTTA GATGGACAGG GGATTACAGT TGATGAGATG GCTTGGATAA TTAGGGGCAT	25140
TGTGAATGCA TTGATTGGTA GATACATAAA ATTAGGTACT TATGCGGCTA AGTATGGTAT	25200
TAGTATGGCA CGCTCGATCT TAAGTAGGGT AGCTGCAACT GCAGCAGCAA GAGTAGGATT	25260
ACTGACCAAG ATTTCTGGAT GGATTTTACG AGTAGCTGTG AATGTAGCTG ATGTATATGG	25320
TAATTTTGGC AACAAATATTG CTGCAGCTTG GGATGCATAT GATAAAATTC CTAACAATGG	25380
TCGTATAAAC TTTTAAATG CGAGAATGAA AGCACTTTGT ATTTTTTTAT TGAATATGTT	25440
AGCTTGGACA GTGCTTGCAA TGATAATTCG TGGAGGGCTA GATGGATTTC ATAGGCATAC	25500
TTGGAGTACT ATTTTAATTG CGTCGCTGTT CGGGGTATAT GATTATAAGC CCATAGATAA	25560
AAATAGAAAA AAGTCCAAAA GAAAAAATAG ATTTGTTTAT GGTAGGGACT TATGAAAGCT	25620
TTACTGACAA AAAAGAAAAC AGTTTACAAA GAAAAATGAT GGAGGAGCAA ACATGGCACA	25680
AAAAGGAGTA AGCCTTATCA AGGCAGCATT TGATACAGAT AACTTTCTCA TGGCTTTTAG	25740
TGAGAAGGTC TTGGACATCG TGACAGCCAA TCTTCTTTT GTCGTCTCTT GTTTACCCAT	25800
CGTGACGATT GGAGTGGCTA AAATCAGCCT CTACGAGACC ATGTTTGAAG TTAAGAAGAG	25860
CAGACGGGTG CCTGTTTTTA AAATCTATCT AAGATCTTTC AAGCAAAATC TGAAACTAGG	25920
TCTTCAGCTG GGTTTAATGG AGTTAGGAAT TGTGTTTCTT ACCCTTTCAG ATCTCTATCT	25980
TTTCTGGGGT CAAACAGCTC TGCCCTTCCA ATTGCTGAAA GCCATTTGTT TAGGTATTCT	26040
GATTTTCTT ACTATCGTGA TGCTGGCTAG TTACCTATC GCGGCACGTT ATGACCTATC	26100
TTGGAAGAA ATTCTTCAA AAGGATTGAT GTTGGCTAGT TTAACTTTC CTGGTTCTT	26160
CCTCATGTTA GCCATTCTTG TCCTCATTGT GATGGTTCTT TATCTGTCCG CCTTCAGTCT	26220
ACTCTTAGGT GGCTCAGTCT TCCTACTTTT TGGGTTTGA CTATTGGTCT TTATCCAGAC	26280
TGGATTGATG GAGAAAATT TCGCAAAATA CCAATAGGAG CTTTATTTCT GAACTACTT	26340
TCAAAGGCTC CAAACGCTAT TCTATAAGCG AGAACTAAA ATCGG	26385

173

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2716 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```

CCTGCCCGCA TTGCCCTAGG CATTAAAGTAA ACATATAAAA GCATGTGAGA GACTGTTGGA      60
AAAGCGAGGA AATTTCCCCT CTTTTCCTCT AGTCTCTCCT TTCTTTTGCT GATTTTATTC      120
AAAGAAAATG ATATAATAGT AGTTATGGAG AAAAAGAAAT TACGCATCAA TATGTTGAGT      180
TCAAGTGAGA AAGTAGCAGG ACAGGGAGTT TCAGGTGCTT ACCGTGAATT AGTTCGTCTT      240
CTTCACCGTG CTGCCAAGGA CCAATTGATT GTTACAGAAA ATCTTCCAAT CGAGGCAGAT      300
GTGACTCACT TTCATACGAT TGATTTTCCC TATTATTTAT CAACCTTCCA AAAGAAACGC      360
TCAGGGAGAA AGATTGGCTA TGTGCATTTC TTGCCAGCTA CACTTGAGGG AAGTTTGAAA      420
ATTCCATTTT TCTTAAAGGG AATTGTGAAA CGCTATGTAT TTTCTTTTA CAACCGGATG      480
GAGCACTTGG TTGTGGTCAA TCCTATGTTT ATTGAGGATT TGGTAGCAGC TGGTATTCCA      540
CGTGAAAAAG TGACCTATAT TCCTAACTTT GTCAACAAGG AAAAATGGCA TCCTCTACCA      600
CAAGAAGAGG TAGTCAGACT GCGCACAGAT CTTGGTCTTA GTGACAATCA GTTTATCGTA      660
GTAGGTGCTG GGCAAGTTCA GAAACGTAAA GGGATTGATG ACTTTATCCG TCTGGCTGAG      720
GAATTGCCTC AGATTACCTT TATCTGGGCT GGTGGCTTCT CTTTGGTGG TATGACAGAT      780
GGTTATGAAC ACTATAAGAA AATTATGGAA AATCCCCCTA AAAATTGAT TTTCCAGGC      840
ATTGTATCGC CAGAGCGGAT GCGCGAATTG TATGCTCTAG CGGATCTTTT CTGTTGCCT      900
AGTTACAATG AGCTCTTCC TATGACTATT TTAGAAGCTG CGAGTTGTGA GGCTCCTATT      960
ATGTTGCGTG ATTTAGATCT CTATAAGGTG ATTTTGAGG GAAATTATCG GGCGACAGCG      1020
GGTAGAGAAG AGATGAAAGA GGCTATTTTG GAATATCAAG CAAATCCTGC TGTCTTAAAA      1080
GATCTCAAAG AAAAGGCTAA GAATATTTCC AGAGAGTATT CTGAAGAGCA TCTGTTACAA      1140
ATCTGGTTGG ACTTTTATGA GAAACAAGCC GCTTTAGGGA GAAAGTAAAA AGTGAGGTAA      1200
TCTATCGGAA TTGGTTTATT TACAGATACC TATTTTCCTC AGGTTTCTGG TGTGCGACC      1260
AGTATTCGAA CCTTGAAAAC AGAACTTGAA AAGCAGGGAC ATGCTGTTTT TATCTTTACG      1320
ACGACAGATA AGGATGTCAA TCGCTACGAA GATTGGCAAA TTATCCGCAT TCCAAGTGTT      1380

```

174

CCTTTCTTTG CTTTAAAGGA TCGTCGCTTT GCCTACCGAG GTTTTAGCAA GGCACCTGAA	1440
ATTGCTAAAC AGTATCAGCT AGATATTATC CATACTCAGA CAGAATTTTC TCTTGGCCTG	1500
TTGGGGATTT GGATTGCGCG TGAATTGAAA ATTCCAGTCA TCCATACCTA TCACACCCAG	1560
TATGAAGACT ATGTCCATTA TATTGCTAAG GGGATGTTGA TCCGGCCGAG TATGGTCAAG	1620
TAICTGGTTA GAGGTTTCCT GCATGATGTG GATGGGGTTA TTTGCCCTAG TGAGATTGTC	1680
CGTGACTTGC TATCTGATTA TAAGGTCAAG GTTGAAAAAC GGGTCATTCC TACTGGGATT	1740
GAATTAGCCA AGTTTGAGCG TCCGGAAATC AAGCAGGAAA ATTTGAAAGA ACTGCGTAGT	1800
AAACTAGGGA TTCAAGATGG TGAAAAGACG TTGCTTAGTC TTTGAGAAT CTCCTATGAA	1860
AAAAATATTC AAGCAGTTTT AGCAGCCTTT CCTGATGTTT TGAAAGAGGA AGACAAGCTT	1920
AAACTGGTAG TAGCTGGGGA TGGCCCTTAT CTGAATGACC TCAAAGAGCA AGCCCAGAAC	1980
CTAGAGATTC AAGACTCAGT CATCTTTACA GGGATGATTG CTCCTAGTGA GACGGCTCTT	2040
TACTATAAAG CGGCGGATTT CTTCATTTTC GCATCGACAA GCGAAACGCA AGGTTTGACC	2100
TACTTGAAAA GCTTAGCCAG TGGAAACCTT GTCATTGCTC ACGGAAATCC TTATTTGAAC	2160
AACCTCATCA GTGATAAAAT GTTTGGAACC TTGTACTATG GAGAACATGA TTTGGCTGGT	2220
GCTATTTTGG AAGCCCTGAT TGCAACACCA GACATGAACG AGCATACCTT ATCAGAGAAA	2280
TTGTATGAGA TTTCAGCTGA GAACTTTGGG AAACGAGTGC ATGAGTTTGA TCTGGATGCC	2340
ATTATTTCAA ATAACCTCCA GAAAGATTG GCTAAAGATG ATACGGTCAG TCAGCGTATC	2400
TTTAAGACAG TTTTGTATCT TCAGCAACAG GTGGTTGCTG TACCTGTAAA AGGATCTAGA	2460
CGCATGTTGA AGGCTTCAAA AACACAGTTG ATCAGTATGA GAGACTATTG GAAAGACCAT	2520
GAAGAATAGA AAGAGGAACA GCTATGAAAA AAACAATTAA TGAGAAGCCG TCGTGATAAA	2580
AAGATTGCGG GTGTTTGTGC TGGGGTGGCC CATTATCTGG ATATGGATCC GACTATCGTT	2640
CAAGTCATTT GGGGTGTTCT TACTTGCTGT TACGGAGCTG GAATTGTAGC TTACATTATT	2700
TTATGGATTA TCGCGA	2716

## (2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13926 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTTTGGTTTT GCCTTATTC AAGACATGAGG GCCATCAGGA ATGATCTGAA ACTGCGAATC

60

175

TGTTAACAGT CTATGGAGAG CTTTCATAGA ACTAAGATTG GGTTCATCTT TGCTGCCACA	120
AATTAGTAAG GTTGGATAAG GGTAAAGTCC TGCTATATCC GTTAAATCAA GTGCTTCAA	180
CTCCTCAGAA ACTCCGACCA TAAGAGTCTT GTCTGCTCCC TGTTTTTCAA ATACTCTTTT	240
GGGAAGTAGT TTAATAATCA GCAATTGAAG ATAAATAGG ATATCCCTG CTAATTTAAG	300
CGGGCATCCT GACAGAATCA AAGCTCGAAG ATTTGGTAAA TCGTAACTGG AAAGTTCTAG	360
TGTCAGGGCA GCACCTAAGG ACAATCCAAT CAAAACAAAA GGTTCGTCTT CTTGAGCTAG	420
GTGCTGATAA ACTCGCTCTT TAGCTTGTTG ATAGTTACTA ACTCCAGAAG GAAATAACTC	480
GATAGCCTCA GAAGGATAAT CTGTCACTAG ATTCCGAAGT TCTTTCCAAG ACTCTGCTGA	540
CTGCCCTAAC CCATGCAAAA ATATTAATTT CATCTAGTTC TCCTCAAGGC TTAATTCATA	600
CAAGCCTCTC ACTGCATTAC AGCCGTAAAT AGCTTCTGCT TGGGTAAAT CTGCCAAGGT	660
CAAGACTTTC TCTTCTACCT GTCCTGTTTC TAGCAAATGC TGACGGTAAA TTCCTGGCAA	720
GATTCCAAGT CGGATAGGCG GTGTGTAGAG TTTCCAGCG ATTTTCAGAA CCAAATTTCC	780
TATAGAGGTT TCAAGCAGTT CTCCTGACTT ATTGTGGTAA ATCTTCTCTT GTTCTCTAG	840
GCTCAAATGC GGTCCGTGAG TGGTTTAAA GTAGGTAAAG GATTGATTCA AAGCAGCTTC	900
CTGAAGACAG ACTTGGCCTT GACAAAAGCT TGTACTGAGA GGGGTTAATA CTTGACGATT	960
GACTTCTATC TCTCCAGATT TGCTAAGGCT GATTGCAAG CGGTAATCTC GATTAGCTTC	1020
ACAATCCTGA CACTCTTCCT CAATCTGTG TCCCAAGTCT TCTGCATCAA AAGGAAAACC	1080
AAAATAACGA CTAGCTTTTC TCAGCCTTTC CAGATGTTGT TCTCAAACA TCAGTTGTTT	1140
TTGGCTGATT TTTCCAGTTG TAATTAATTG GAAGCGAGCT TGTTTACGAT AGAGAACTGC	1200
TGCCTTTTGA TGAACCTCTC GGTATTCAGA TTCCCATGTG CTATCCCAAG TAATCCCTCC	1260
GCCAACTCCA TAAATGGCTT GACCTTTGTG AAGTTGAATG GTACGAATGG CCACATTAAA	1320
AATCCGTCGT CCATTTGGAA GCAAGAGACC AATCGTTCCA CAGTAGACTC CACGCGTTG	1380
AGGCTCCAAG TCCTTGATAA TCTCCATTGT CGCAATTTTC GGTGCACCCG TTATGGAACC	1440
ACAAGGAAAG AGTGAGCGGA AGATTTCAAC AAGGTCCACA TCCTCTCGCA ACTGACTCTT	1500
GATGGTCGAA GTCATCTGCC AAACAGTTGA ATACTGCTCT ACCTGACACA GACGCTCCAC	1560
GTGCTCGCTC CCAACTTCAG AAATACGCTT CATATCATTG CGCAAGAGGT CCACAATCAT	1620
CATATTTTCA GAGCGATTTT TGGGATCCTG TTCCAACCAA CTGGCCTGTT CAAGATCTTC	1680
TTGGTCAGTT ACCCCACGCT GAGTCGTCCC CTTCAATGGT CGTGTGTCA ACTCGCGATC	1740
ATTTTGCTCA AAAAAGAGCT CTGGGCTCAT GGAAATCACT GTCATCTCGT CATGTTCCAC	1800

176

ATAGGCATTG TAGCCCGCCT CCTGCTCTAC CACCATACGA TTGTAGATGG CAAAAGGATT	1860
GGCATTTAAC TTTTGCTTAA GTTGGACGGT GTAGTTGACC TGATAGGTAT CTCCCTGCCG	1920
TAAATGATGG TGAATTTGGG CAATGGCCTT TTCATAGTCT GCTGCAGACG TTACTTCCTG	1980
CCAATTTGAG GGCAAATCAA TATCCTCATA AGTCAGAGGA ATAGGGGAAG TTTCTACGAT	2040
ATCATGAACA GTAAAGTAAA GCAGGTACTC TCCCAGTAGG GGATCCTTGT GAACTGCTAA	2100
TTTTTCCTCA AAAGCAGGTG CAGCCTCGTA GCTGACATAC CCCACCACAT AATAACCTTG	2160
CTCTTGGTAG CTTTCCACTT GTGCCAGCAA ATCTGCCACT TCTTCTACAT TTCTCGTTTT	2220
CAACTCTTTA ATAGGCTGGG TAAAGGTATA TCTCTCCCC AAAGTCCTAA AATCAATCAC	2280
TGTTTTTCTA TGCATACCTT AAGTATAGCA TAAAATAAGA AAACCCTCAT CCGCAAAGCA	2340
GATGAGAGAT TTCAATTATT TAAAGATTGA AGTTTTAAAG CTATTTGTTT GTTGAAGAAG	2400
TTTCTTATAA ACAGCTTCTT TTAATTTAAC TGTATTATTC ATAGATACTG TTTTATTACC	2460
GTTTGCTTCT TGTTTAAGAG TTTCCGGATC TTTTTTAAAC GCTTCTTTAA ACAATGTCAG	2520
TAAATCATCG TATGATGAAA CGGAAGAACC ATTTACTTCG AATGTTGTTA ATCCTTTCGT	2580
TGCTTTATCT TTAACCTCTT TGAAGTAAGC TTTTTTAAAT TCTTCAATAG TATTAAATGT	2640
ATTGTTAGAT ATTTTCTTGA TAATATATTC ATCACTTAGA ACAGACTCAC CATCTGTTTT	2700
AGATTGTTGT TTATATTTAT TTGAAGCATA ACCTAAGAAC CCATTTTCGT ATCCGTAGTA	2760
ACCCCATAA CTAAAAGCAT TATGTTTGAA TGAACAGCT CCAGGAGCAC CTTTACTAGT	2820
ATTACCTCCG TAGATACCGG TCATCATCTT AACACCTACA TAAGGTGATT GATCGTTATA	2880
GCTAATTGCT TCGGGTTTAT AGATACCATT ACCTGGATTG CGATTAGTCA TTAATTGTTG	2940
ATCAACTAAA TCATTAACAG ATTGAATATT TAATTCATTT TTCTCTCTT GACTTAGATT	3000
TCGAATTTTA TCCCATTGAT TTAATTTATT GTTATCACGG TATTCTCTAT CTATTTTTTT	3060
GAACCATGCA CTATTTAAAT CTTTATTTTG TTGAGAAATC ACAGATTGAG CCTCAATTTT	3120
ATCAAGAAGA GTTAAAGTGT CATTATAACC CTTTATATAT CTATTAATAT CTTCTCGTGT	3180
TTTTAGAGTT TTTGGATCTG TAATATACCA CTGATCCCA TCATTTTTCG GTTTAAATAC	3240
CATATTAATA CCTAAAGAAC CAACTCATC AAATCCACTA CCAGTAACAG GAGTTGTAG	3300
CATACCCTGA GCATATGCTT CAGCATCAGT ACCTTCACGG TGTCCAAAGC CACCTAAGTA	3360
AATCGCACGG TCGTTGACGT GTGTTGTTTC ATGTGTGTAA ACTGAAATAC CGTATTCACC	3420
AACCATTTCT AAATGAACAT ATTTTACATC AGTTCTAATA TCATCAGAGT TAGGATATAT	3480
AGCAGCATAA GCTCCTGTTT CATTATAATT ATAATACTTA TCCATAGGAC CAAAGAATTC	3540
TCTAAGAGGA GTATATACTT TGTCCGTATT ATAGCGGCCA TATTTTCAA CCCATCCACC	3600

177

AGGAGCGTTA TAACCTTCCC AAATAGGAAT AACAGCATCT CTTAGTAGTC GTTGTTTAAC	3660
GTTATCAGAC GCTAGACGAT ACCAGAAATC ATAATAGTTT CTATAACCAT CTGCAGCTTT	3720
GTTAACGATA TCTTTAATAT CTTCTAATGA TTTTTTACCT AATCGCTCTG CACTACCAAA	3780
GGCAATTGCA TTATAATTTG AAATTAAATA AAGATGTGCT TTATCAATAT TCAGTAGTGG	3840
GAGTATAGTA TTTCTAAGGT GACTTCGTTT TAAATTATCG AATGCACGAT GTTTAGAATT	3900
TTTAATTTCT TCGACCTCAG AAGCGCGTTC TGCGATGTAG ACATGGTCTT CTGTAGCATC	3960
AATAAACCAA TCGTTCATAT TGTCTATATT TGTGAACAAT TGTCTATTAT AATTAAAAA	4020
TGCATCTAAA TTACCTGATT TAGTATATTT AGCCAATACT TGACCGAATG CGTCCAATGT	4080
ACGTGAACCT TTAATGTTGT TCTCTTTAGA ACCGATTTC AATTAATCTGT CTAATACGCT	4140
AACTTTTTCA CCATAGAAAT CTGGTTTGAA TAGCATTAAAT TCTTTAATAT TAACATCACC	4200
AAATTAACT CCATAGTAAC GATTTAGGTA AGTTAAACCT AGTAATAAAG CTGCTTTGTT	4260
TTTCTCGACT TTATCACCAG TCATTTGACG AGCAGCTGCA GAATCATTTA GTTGATGTTT	4320
TTGTTTTTGA ACTAATTTTG TGATTAGGTT TGTTAAGTTT TCTTTAACAT CTGTGAAGCT	4380
TTCTTCTAAA TATAAATCTT TGATTGCATT AACTCTATAG TCACCTAATC GATTTAGATG	4440
CTGATACATC GTTTGAGACT GAAGCTCTAC TGATTCTAAA ATAGATTTTA TATCATTAA	4500
AAGAGTAGTG TTATCTTTT GAACGATATT AGGTGTATAT TTAATTCCTA AGTCAGTTAT	4560
AGTATATTCT TTTACATTAC TTAAACCTTC ACTGCTAGAA GACAAGTTAA AGTAATCTTT	4620
TGTACCGTCC GCATAGTGAA CAATAATTTT ATTAGCTTCA TCTAGGTTTG TGATAAACTC	4680
ATTGTTGTTT ATCGCGGTAA CAGAAAGAAC TTCTTTAGTA TTTAGATGGT GTTCTTTATT	4740
TAATTTATTA CCTTGATATA CAATATAATC TTTATTGTAG AATGGTATTA ATTTTCAAG	4800
ATTTTATAG GCTTGGTTAT ATTCAGCGTT ATAATCTTGA ATACTAGAAT AGGCTTTTTC	4860
TTCATTAAGT TTTGCAAGAG GAGATAGATC ACTTCTAAT TTATCAGCAG TAATATTGAA	4920
AGTAGTAACT TTAGCATCAG CTGTTCTTT AGTTAATTTA GTAAATGTTT TAGATTTCTT	4980
AAATGATCTA TTACCTGACG AATATCCCTC TACCGCATAT AAATCTTTTA TATGAGCACT	5040
AGCATAATCA GAATCATCAA CGTCGTTAGA GCCGAATAAC TCCTCTCCAC GGATAATCTT	5100
AGCATAGCTG ACAGAATTAC TTACCGTACC TACAGGCCAA GTCTTACTTG CTATTGCTCC	5160
AACTTCTACT GGATTGAAA CATCTATTTT ACCTTTTACA ACCGACTCAG TTAGGAGAGC	5220
TTTTGTACCA ATAAGATGGT CTAGAGTTAA TCCATAATCT ACTTTAGGAA CTAACAAGCT	5280
GGCGCGTGTT TTGTTTCCTG TAATAGTAGC ATCAACATAT GCTTTTCTAA CAATTCCTCT	5340



178

ATAGTTTGTA CCTGCAATTC CCCCTGTATG AGAGCCATTT CCACTTGTAG AGTGTAGTTT	5400
GCCAAAGAAA GCAACATTTT CAATACGAGT TCCATCATTC ATATTATTTA CAAATCCAGC	5460
AACATTATTA CGACCTGAAA GTGTGCCTGT AATTTTGACA TTTGTAATAA CTGAAGAACC	5520
TTTCATAGTA TTGGCTAATG ATGCAATATT ATCTTGACCA GAACGTTCTA TCTCTACATT	5580
TTCAAAATTC ACATTATTTA TCGTTGCGTT TGTATCACA TTAAATAATG GATGTTCCAA	5640
TTCAGTAATA GCAAATTGTT TTCCTTCAGA ACTTAAAAGT TTTCTGTGA ATTCTTTAGT	5700
GATATATGAT TTTCCATTAG GAACAACATT TCTAGCGCTC ATTGATTGTC CCAGACGATA	5760
TTCTTTTGAA GGATCGTTTT GAATAGCTTC CACTAATTCT TTGAAATTAT AATATACATT	5820
ATCTTCGTGG ACTTTAGGTT TTTCAATATA GTGAACGTAT TCTTCTTCAA ATTTATATC	5880
AGCAGTTCTA GAGACTAAAT TGTCTGCGAT TGTGTAACT TTATATACAG GTGTTCCGTT	5940
AACCGTAGTT TCTTCTATAT TTTTAACAGC TAGTAATGTA GTTTTCTGAT TATTTGAAGT	6000
TATTTTAAAA TAATAATTGC TCTTATCATC AGGAATAGTT GTTATCAGTG ATTCATTAGT	6060
TTCTTTTCCA TTTTCGTATT TGATTAAATC TGTACGTTTA ATATTTTAA GCTCAACTTT	6120
TTTAAGATCT AATTGAATAT TTTGATTTTC TAGAGTTTCA GTTCTTTCAC CGTTACCTCT	6180
GTGTAATC ATAGTTGTAG ATAGGGTGTA TCTTTGTAG TACTCTAGGT TCTTAAATGC	6240
AGCGCTTATA GTTTCTGTTG TTACCTTGTC ATCTGTAAGG ACTACAGTAT TAATAACTTC	6300
TTCTCCTTTT TTCAATTCAG CTGTGATTGA TTTGATTTT GTTTTGTTT GATTTTCTAG	6360
AGTATACCTA GCAACAGCTT CACGTTCCAA TATTTTCTTA TCGGTACTAG TCAATGTAA	6420
TATTGGCTTT TCAGATAATT CAACCAATTT TTCAATAGTT GCAGTTAATT TTTCAACAGC	6480
TTGTTAACT TCACTTTGTT TAGCATCTGT ATTAGCTGCA ACTTTTTCAG CCTTTGTAAC	6540
TTCAGTTTGG AGGTTTGGC AACTTCTATC ACTGTAATGT TCTTTTACCT TTGTTTGTGC	6600
ATCTGCAATC GTATTGTTTA ATTCAGTTT ATCAACGTTT AGAGCGTCAA TAGCCGTTTT	6660
AAGTTTATTT GTCTCGCTAT TTACCTCAGG CTGTTTACA GGCTCTGAAG CATAGACACC	6720
TTTTGCAGTT TCTAAACAG GTCCAAGAGC ATTGTAACCT GCTGTAGAA ATCAGTAGG	6780
AGAAACTGAA CTAGCTTTAT CAATTTGATT ATTTAACTCA CTTTATCAA CTGGTTCTTT	6840
AGTACCAATA CCTTTATTT TATCTTCTGG TTTGCGTGT TCCCTACAG CCTTCTCTC	6900
TTCAAGAACT TCTGGTTGCT TTTCTGGCTC AACTGGTGCC GTTGGTGCCT GTTCGTCTC	6960
TCTTGGCGCG ACTGGTTCAC CTGCTTGTTC AACTTTTGGT TCCCTGTTG GTTCTGTTG	7020
TTTTTCTACA GCAGGCGTTT CAACTTTTGG TTGTTCAATA GATTGATTAA CAGTCTCCTC	7080
TTTTGGTTCT ACAGTTTCTT CAGCCTTGGT ATCTGGAGTT GACTCTTCTT GTTTCGGTGT	7140

TTCTCTACA GCCTTCTCTT CTTCAGGAGC TTCTGGTTGC TTTTCTGGCT CGACTGGTGC	7200
CTTTTCGTCT TCTCTGGCG CGACTGGTTC ACCTGCTTGT TCAACTTTTG ATTCTCAGC	7260
TGGTTTGTCT GATGGTTGAC TTTCTGGCTT AACTGCTACT TTTTCTCTG GTTTTGACTC	7320
AACTTCTCCA CCTACTTCTT CAACTGGAGC TGGTTCTGCT GAATCTTCTT TCCCCTCTTC	7380
TACTTTAGGA AGGGTGTCTT CAGTAGGTTT TACCTCCGAT TTTGGTTCCTT CCTTTGACT	7440
TTCTTCTGTT TTAGGTGCTT CTTCTTTTGG AGCTTCCTCT GTCTCTACTA CTTGGTTTTT	7500
TGTCCTAGCT TGCTCCTGAT TTGTTATTGA TTGAGGAGTC TCAACTTCGA CCACAGTCAC	7560
CTCTCCAGGT TTTGCTGAGG TTTCTTCTAA AACAGTGTCC AAGCCAAGCG TTTTGAGGAT	7620
GTCACCTGAT AGATAACCAA CATAGCGATA GCCCTCCATT TCAACAACAC CCTCTCGACT	7680
AGCCAGCGCT AGGGTCGCAA CTGGGTCTAC AGCCCCTGCA CTAGGAAGAA CTACCAATCC	7740
CATAGCTCCA ACTAGAAAGA CGCTAGCAAT TTTCTTTCTC TTGTAGATTA AAAGCAAGCT	7800
CCCAACAGTC AGCAAACCAA AAGCTGTCAA AACAGATGCT TCTGTCCCTG TTTGAGGCAA	7860
CTGATCTTTT TGATACACCA AACCATATAC AACTTCATTC CTGTCAGGCT TTCCTGTCTG	7920
AATTAAATCT TTAGCTTCTT GTGAAATAAT CTCTTTATTT ACATAGTGAT AGGTGGCTGC	7980
GTCCACTACA GAAGGAGCCA TCAAAAGGCT TCCAAGAAAT ACAGAGCCTA CAACTCCCTT	8040
AATCTTACGA ATTGAAAAAC GGTCTTTTTT AAACACTTTT ATCTCCTTTA TTCATTCTCA	8100
AAACTTCCTA ATAGCATCTT GCGGATAGTG CGCAGCGCA CCTCCGATTA ATTTTGGACG	8160
ACTAGCCAST GCCCTTACAT GGGCATGACC AATCTCTCTC AAAATAGGGC GAATCGGAAC	8220
CTGAACATGC TTGACATGCA TGCCAATTGC AGTGTCTCCG ATATCCAATC CAGCATGAGC	8280
CTTGATAAAT TCAACCTCAA CTGGATCCTG CATAAACTTA AAGGCTGCCA ACTGCCCCGA	8340
ACCTCCTGCA TGAAGAGTAG GATGGACACT GACAATTTCC AGACCAAAT GCTCTGCCAC	8400
CTGACGTTC AACAAGAGAG CCCGATTGAC ATGCTCACAA CCTTGAACCTG CTAAATGGAT	8460
ACCTCTACTA CCTAGAATAT CCAAGATAGT CTCCACTATC AGCTCACCAA TCTCTTGACT	8520
GGATTCTTTC CCAATATGAC CACCTAGCAC CTCACTAGAA GATAGACCTA AAACAAAAAG	8580
GGCCCCCTGC TTCAAATTGG TCTTTTCTAA AACATCTTCC ACTACCTGAC GTGTTTCTCT	8640
TTGAATCTGT GTCTCGTTCA TCTCTGTAC CTCTGTTGTC ACTCTTCTAT CATACCGTTT	8700
TTTCTTGTTT TTAGCAAGAT AGACAACCTA GAAAGTTTGC CCAATTACGC ATAAATCTCC	8760
CAGAAATGAC TGGGAGTTAG CTAGTTTCTA TTCTATTTAT ATATATTTCA ACTTTCGTCC	8820
CTTTTGGGG TCTAGAATCA ATCTTCATAT GGTAATTGGC TCCAAAATGA AGTTTGAGCC	8880

180

GTTGATCGAC	ATTTTGAAGA	CCAACTCCCC	CACGTTTGAG	TTGAGTTTGA	CTACTATCAC	8940
CAGCATCTTG	GAAGCCAACG	CCATCATCCT	CAATACGGAT	GACCAATCCC	GAATCCTGTT	9000
TCTGGACAGA	AAGTTTAATA	TGGCCCTGAC	CTTCCTTTTC	CTTAATGCCA	TGGTAAAGAG	9060
CATTTTCTAC	AAGGGGTGT	AGGACCAGCT	TGGGTAAGAC	TAAATTATCA	AAGGCAACAT	9120
TTTCATTAAT	TCGTATTC	AGCTTATCTC	CATAGCGTTG	TTTCTGGATA	AAGAGATACT	9180
GGCGGACATG	ATTGATTTCG	TCAGAGAGAC	AAATCAAGTC	CTTGCCTTGA	TTGAGCGCCA	9240
AGCGGAAATA	GTTTGCCAAG	GACTTGGTCA	CCTGCACCAC	TCGCTGACTA	TCATGAAAT	9300
CAGCCATCCA	GATGATGGTG	TCCAAAGTGT	TATAGAGGAA	ATGTGGATTA	ATCTGGCTCG	9360
AAAGGGCTTG	AAGTTGGTAC	TGACGGGTCG	TTTCTTCCTG	GCTACGAATA	GCTACCATCA	9420
ACTGATCAAT	CTGATCCAAC	ATAGCATTA	ATTGGCGAGT	TACTTCTCTC	AGTTCATAGG	9480
CACCAACTTC	CTTGGCACGA	AGATTTTGAG	CACCAGAAGC	AATTTCCAAC	ATGGTTTCTC	9540
TCAAATCCTT	CAAAGGAGCA	ATCCAGCGTT	TAAGACTGAA	CCACACTAAG	CAGAGACAGA	9600
CAAGAAGAGA	TGTGACACTG	GCCCCAAGCA	AGGTCCACAA	GAGCTGACTC	CGAACCTGGT	9660
CTAACTTTTC	CAATGATGAC	ACGCCAAGCA	CCGTCCAATC	AGTTCCTGCA	ATCTTCTCTT	9720
GACTGACGTA	GGATTTGTGA	CCAGGAGTAT	AACCCTGACC	TGTATCGATG	TAGGGTTTCA	9780
TAGCCTCCAT	TTTGCTAGAC	GAECTATAAA	CTGTGTGTTG	AGGATGGTAG	ACAAATTCAT	9840
GGTTTTTATT	GATAATGAAG	GCAAAGCCCT	GCTGCCCCAA	CTGGAGTTGA	TTGAGATAGG	9900
CTTCCAGAGT	TTCATAAGAA	ATATCCAAAC	GAAGCACACC	AAGATTGGCT	CCCTTTGCAT	9960
CAACAAGTTC	TTGAGTGACA	GAAATGACCC	ACTGACTATC	TGATTTACGA	GCTGGAGTCA	10020
AAACAGGCAT	AGCTCCCTGA	TGAATGGCCT	TTTGGTACCA	ATCCTCAGCC	ATCATATCAG	10080
AGGAAGTTTT	CATCTGCACA	CTGTCATCTG	TAGAAATGAC	CTGACCAGAT	TTGGTCACCA	10140
GCACAACAGT	TTTCAAGTCC	TTATCTGACT	TCAAGATGGT	CAAAAACAAA	TCTCGGATTC	10200
CCTCGACCTT	GTCTTGACTG	GGATTCTCAG	CATAGGCCAG	AACATCCGTC	TGCTGGGTCA	10260
AACCAGTCGA	GGTGGTTTCT	AGTTTTTTGA	TATAAGACTG	AATAAAGTGG	CTAGTCTGGC	10320
TGATGGTCGT	TTGGCTGTTG	CCCTCAATGG	TGGCCTCAAT	GGCTGAAGAA	CTTGATTGAT	10380
AGTAGAAAGT	TCCAACCAGA	GCTAGGAGAA	TGAGAAAGAC	CAGAAAGATG	GAAATAACCA	10440
TTCTAACTAA	AAGAGAAGAA	CGCTTCATCG	GTCTTCTCCC	TTCTTAAACT	GACGAGGTGT	10500
CACACCTGCA	ATCTGCTTAA	AACGTTGGGT	AAAATAGTTC	ATATCTTCAA	AACCAACCTT	10560
CTCTGCGATC	TCATAAATCT	TCAGATCTGT	AGTAAAAAGC	AAGAGCTTGG	CTTGTTTAAAC	10620
ACGTTCTCTC	ACCAGATAAT	CCTGAAAAGG	CAAGCCCAAC	TCTTTCTTAA	TCAAGGAACT	10680

181

CAGATAGGTC	GGA	CTAA	CTAAGTCACT	GGCTAAAGAC	TTTAACTAA	ATTGGCTATC	10740
AGCCAGATGA	GA	CTGGG	TCTGGGCCAT	GTTTCCTTCA	AACCTATTAG	TCAATAAATC	10800
TTGTAAGTGC	CTCTCTTCT	CTTCCTTGTC	TAGTTTTTGT	TTGATTTTCC	CCAACATTTT		10860
CTCAATATCC	TGACGAGAAA	AGGGTTTGAG	CAGGTAGTGG	TCCACACCTA	GTTTGACAGC		10920
AGACAAGGCA	TAATCAAAAT	CATCGTAACC	TGTTAAAAAG	ACCAAATGAA	CCTGAGGATA		10980
GGTTTCTCGT	ACCAGACTGG	CCA	ACTGGAT	GCCATTTAGA	TGAGGCATGT	TGATATCGGT	11040
TAAAATGATA	TCTGGCACCT	GCTTTTGAT	CAATTC	CCCA	GCCTGCCTTC	CATTTTCAGC	11100
CTGACCGATG	ATTTCCATAT	CGTAGGCTGC	TACATTGACC	AGTTTAGTCA	AACCTTGTCT		11160
TACCAGATAT	TCATCTTCTA	CGATTAAGAT	TGTGTAGGTC	ATGCTCTGCT	CCTTTACCAC		11220
TTACTAGTAT	CAGTATAGCA	AAATTCCTCT	CTAACTGCTT	AGGAAAGACC	TCTTATACTC		11280
AATAAAATC	AAAAAGTAAA	CTAGGAAGAT	AGCCACAGGT	TTCTCAAAGT	ACCGCTTTGA		11340
GGTTGTAAAT	AAA	ACTGACG	AAGTCGACTC	AAAGTATAGC	TTTGAGGTTG	TAGATAAAAC	11400
TGACGAAGTC	GATAACCTA	CATACGGTAA	GGCGACGCTG	ACGTGGTTTG	AAGAGATTTT		11460
CGAAGAGTAT	TAATCAACAT	AATCTAGTAA	ATAAGCGTAC	CTTTTCTTC	CATTTGGTCT		11520
TTGGGAATAA	AGCGGATAGA	GAGGCTATTG	ATACAGTAAC	GTAAGCCGCC	CTTGTCCTGT		11580
GGACCATCCG	TAAAGACATG	CCCAAGGTGA	GAATCTCCTA	CTCGGCTCCG	CAC	TTC	11640
CGCGTCATAT	TGTAGGACTT	ATCTTCCTTG	TAGGTGACAA	CATCTGGACT	GATGGGTTGG		11700
GTAA	AACTAG	GCCAGCCACA	ACCAGACTCA	AATTTGTCTT	TTGATGAAAA	GAGAGGTTCC	11760
CCAGTTGCTA	TATCCACATA	GATACCGGAT	TCAAATTTAT	CCCAGTAACG	GTTTGAGAAA		11820
GCTCGTTCTG	TTTGATTTTC	CTGGGTAACT	GCATACTCCT	CAGGTGACAG	GGTCTTTTTC		11880
AATTCCTCAT	CAC	TGGTTT	TGGATATTTG	CTGGCATCAA	TGACAGGATA	GGCCGCCTGA	11940
TTAACATTGA	TATGGCAGTA	GCCATTTGGA	TTTTTCTTGA	GATAGTCTTG	ATGGTAATCC		12000
TCAGCCACCA	CAAAATCTT	CAAGTTTTC	TTTCAACTG	CTAGAGGTTG	ATCGTATTTT		12060
TTAGCCACCT	CATCAAAGAC	TTGGTTAATC	ACTTCCAAAT	CCTTGTCATC	TGTGTAATAA		12120
ACACCAGTAC	GGTACTGGGT	CCCCACATCA	TTTCCTTGTT	TATTTTGTCT	GGTTGGATTG		12180
ATAATGCGGA	AATAGTGAAG	CAGGATTTCC	TTGAGAGAAA	TTTGCTTGGC	ATCATAGGTG		12240
ACATGGACGG	TTTCTGCATG	ACCTGTTTGG	TTAATCAATT	CGTACTTGGT	TGTTTCTCCT		12300
CTACCATTTG	CATAGCCTGA	AACGGCATCC	GTCACCCCGG	GAACACGTGA	GAAATATTCC		12360
TCCACTCCCC	AGAAACAACC	TCCAGCTAGA	TAAATTTCTG	GCAAGTCTGC	GTCTTTACTA		12420

182

ATTTCTGTTT TTTTCACTGC TTTTCCTCCT TGGCTAACTG CCGCCTTTTC AATTTGCGAG 12480  
 GCATCTGTCT GCCCTGCATT TCGTATCAAT AGAACATAGA AACCGGTTAT GGCTAGAAAA 12540  
 AATACTCCTA GCAACAAGAA GATTTTAAAC TTATCATTCA TAAGACGCCT CCTAGGCTAA 12600  
 TTCCTTCAAA GTTTGCAAAA TTGCATCTTT TTCCATGAAT CCTGGATGTG TTTTGACCAG 12660  
 CTTGCCTTCT TTGTCTATAA AGGCTTGGGT TGGGTAAGAA CGGACACCAT AAGTTTCCAA 12720  
 AAGTTTGCCT GATGGGTCAA CTAGGACTGG GAGATTTTAA TAATCCAATC CTTTATACCA 12780  
 ATTCTTAAAG TCCGCTTCAG ATTGCTCTCC CTTATGTCCT GGTGACACTA CTGTCAAGAC 12840  
 CACATAGTCA TCACCAGCTT CTTTAGCAAT CTCATCCGTA TCTGGAAGAC TAGCCAGACA 12900  
 GATGGAACAC CAAGAAGCCC AGAATTTGAG ATAGACTTTC TTGCCCTTGT AATCAGATAA 12960  
 ACGGTAGGTC TTGCCATCTA CTCCCATCAA TTCAAAATCA GCCACCTCTT TCCCTTTAGC 13020  
 TGGCGTTGTT TTACTAGCTG TCTGCTCCGT CTTCAATTCA TCTTTCGTTT GGTGTTCACT 13080  
 AGTCACGGAC TTGCCTGAAC AAGCCGTCAA ACAAGGAGC GAACCTGCTC CAAGAACACA 13140  
 TGTITGCCAT TTTTTCATAT TGATATTCCT TTCCATTTTA TTCAAATAAT TGACTTAAAA 13200  
 TTGAAGCATT TCCAAACAGA ACCAAGAAGC CCATCACAAT AATGAGAAAA CCACCCACTT 13260  
 TTTTGAGGAT TCCGAGATAG GGATGAAGTT TTCGGAAATG TTTCAAAACA TAACTAGAGG 13320  
 TCAGAGCTAG AAGCAAGAAT GGTAGCGCCA AGCCAGCGT ATACACCAAC ATGAGACCAG 13380  
 CTCCCTGCCA AGCTCCTGAA CCACCTGAAG CCGCCAAGGC CAAAACAGAC CCCAGAACCG 13440  
 GCCCCACGCA AGGCGTCCAA GCAAACTAA AGGTCAAGCC CAATAAAAAT GCCTGACTAT 13500  
 AGCCCTTACC ATTTTGCCCC TGTCTTGCA GTTGTAGCCT CTTTTCCTTA TAAAGCCCCT 13560  
 TAAAGTGTAG AATCTCCATT TGGTGCAAAC CAAGAAGGAT AATAATTGCC CCAGTAAGAT 13620  
 ATTGGAACCA AGAAGCATAA AGCAAATCGC CTAAAAACC AGCTCCATAG CCCAACAAAA 13680  
 TAAATATAAA GGAAATTCCT GCTATAAAGG CCAGAGTTCG TAATAAACTA GTAAGTGA 13740  
 TTGAAAATTT GCCGCTAGAA GCCTGAGCAC CATCCTTATC ATCTAGTAAC ACTCCTGTAT 13800  
 AGACCGGTAA CAAAGGTAAG ATACAAGGAG AAAAGAAGGA TAGAATCCCT GCCAAAAAGA 13860  
 CACTTAGAAA AAAGAAAATA TGACCCATAA AGTTCCCTCT ATCATTTTAT TGATAGATT 13920  
 ATTATA 13926

(2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20199 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CCCAGCAGAA AAATGGCATT TGGAGATAAT GGAAATCGTA AAAAACTAT GTTTGAGAAA	60
ATAACCTTGT TTATCGTGAT TATCATGCTA GTAGCAAGTT TATTGGGAAT TTTTGCAACT	120
GCAATTGGTG CCCTCAGTAA TCTATAAAAT AGATTCAAGA AAATTTAGTG ACTGGGATTT	180
CCCAGCCCTT TTTTAAAGTG AGAAGAAATA ATGAGTATGT TTTTAGATAC AGCTAAGATT	240
AAGGTCAAGG CTGGTAATGG TGGCGATGGT ATGCTGCCT TTCGTCGTGA AAAATATGTC	300
CCTAATGGAG GCCCTTGGGG TGGTGATGGT GGTCTGGAG GCAATGTGGT CTTCGTTGTA	360
GACGAAGGAC TACGTACCTT GATGGATTTC CCTACAATC GTCATTTCAA GGCTGATTCT	420
GGTGAAAAAG GGATGACCAA AGGGATGCAT GGTCTGGTG CTGAGGACCT TAGAGTTCGA	480
GTACCACAAG GTACGACTGT TCGTGATGCG GAGACTGGCA AGGTCTTAAC AGATTTGATT	540
GAACATGGGC AAGAATTTAT CGTTGCCAC GGTGGTCGTG GTGGACGTGG AAATATTCGT	600
TTCCGCGACAC CAAAAATCC TGCACCGAA ATCTCTGAAA ATGGAGAACC AGGTCAGGAA	660
CGTGAGTTAC AATTGGAACT AAAAATCTTG GCAGATGTCG GTTTAGTAGG ATTCCCACT	720
GTAGGGAAGT CAACACTTTT AAGTGTTATT ACCTCAGCTA AGCCTAAAAT TGGTGCCTAC	780
CACTTTACCA CTATTGTACC AAATTTAGGT ATGGTTCGCA CCCAATCAGG TGAATCCTTT	840
GCAGTAGCCG ACTTGCCAGG TTTGATTGAA GGGGCTAGTC AAGGTGTTGG TTTGGGAACT	900
CAGTTCTCCG GTCACATCGA GCGTACACGT GTTATCCTTC ACATCATTGA TATGTCAGCT	960
AGCGAGGGCC GTGATCCATA TGAGGACTAC CTAGCTATCA ATAAAGAGCT GGAGTCTTAC	1020
AATCTTCGCC TCATGGAGCG TCCACAGATT ATTGTAGCTA ATAAGATGGA CATGCCTGAG	1080
AGTCAGGAAA ATCTTGAAGA CTTTAAGAAA AAATTGGCTG AAAATTATGA TGAATTTGAA	1140
GAGTTACCAG CTATCTTCCC AATTCTGGA TTGACCAAGC AAGGTCTGGC AACACTTTTA	1200
GATGCTACAG CTGAATTGTT AGACAAGACA CCAGAATTTT TGCTCTACGA CGAGTCCGAT	1260
ATGGAAGAAG AAGCTTACTA TGGATTTGAC GAAGAAGAAA AAGCCTTTGA AATTAGTCGT	1320
GATGACGATG CGACATGGGT ACTTTCTGGT GAAAACTCA TGAAACTCTT TAATATGACC	1380
AACTTTGATC GTGATGAATC TGTCAAGAAA TTTGCCCGTC AGCTTCGTGG TATGGGGGTT	1440
GATGAAGCCC TTCGTGCGCG TGGAGCTAAA GATGGGGATT TGGTCCGCAT TGGTAAATTT	1500
GAGTTTGAAT TTGTAGACTA GGAGACTGGT ATGGGAGATA AACCGATATC TTTCCGAGAT	1560
GCGGATGGTA ATTTTGTTC CGCCGACAC GTTTGGAATG AAAAGAAATT GGAAGAATA	1620

184

TTTAATCGTC TCAATCCAAA TCGTGCCTTG AGATTGGCAC GAACTAAAAA GGAAATCCA	1680
TCTCAGTAAA GAAGCTAAAA AATCCCGTGC CTCATCAGAC ACGGGATTTT GTGGTACGAC	1740
AGGCATGTAT AGCAAACTGA ATCTGGAATA GCACAGCATA TCTTCTAAAA TATAGTAAAA	1800
TGAAATGAGA ACAGGACAAA TCGATCAGGA CAGTAAATC GATTCTAAC AATGTTTAT	1860
AAGCAGAGAT GTACTATTCT AGTTTCAATC AACTATATTG TTATAAATTG ATTTGAATTT	1920
CAAAATTAAA TTGTTTGATT CTTATTTCOA TTTGTTATAG TATATCTGAT GTCAAACTTC	1980
TCGGCGAGTC AAATAGCGAT TCCCAAGCCT GACTATCGTG AGGTAGCGGA TTAATATGGT	2040
CTGGGGATAG ACCGTTTAA GTCTGACGCT GGAAATAAGA ATTGTCAGAA GAAGGGATAG	2100
CGAAATCGTG GCTCTACGAA CAGGAACGTG ATAATAAGGC GTATATAGCG GATPAGAGGG	2160
CATCAAACTC TAAAGTCCAA AAAGGTAGTC GTAACCTATA TGGCTAAATC ACGAGAGTAA	2220
TTGAATTCGT ACTAAGATTT TCTATTTTCA CTGTAACCTT TTAACGCCCT TATATCTTGT	2280
ATACACGAGG AAAGATGTAC GACTTATCCC GTGAGGTCTA TCACTATAAA GAGAAAACGA	2340
CAGATAGAAG TGATCCTGAG TCACGGTTAT CTGCTGATA GGACGGTATG TATAAACGC	2400
TTCTGTGAAC TGAGAGAAGG GGGAGAAGTT CTTGCTAAAA TTTAGTTGAA CAGCCGTATT	2460
CCGATACTTA GATAAGAGAT CTAGTCTTAG CTCCTACTCA GTTTTAGGGG ATAAAAAAGG	2520
GGCAATAGCG ATTCGAGAAA GATTATACTC TTCGAAAATC TCTTCAAATC ACGTCAATAT	2580
CGCCTTGTCG TATGTGTAGG ATACTGACTA CGTCAGTTCC ATCTACAACC TCAAAACAGT	2640
GTTTTGAGCA ACCTGCGGCT AGTTTCCTAG TTTGATCTTT GATTTTCATT GAGTATTAGT	2700
AATTCAGTTA CTAACCTGTC AACTCTGATT TATCCAATAA AATTGAAAAG GATGGAAAAA	2760
AGGATAAATT TATGATATAC TTTATTTTGA AGACCTTATT AGAAATCTTG AAAGAGTATT	2820
GAAAACCTAG AATGAGAAAA ATTGTTATCA ATGGTGGATT ACCACTGCAA GGTGAAATCA	2880
CTATTAGTGG TGCTAAAAAT AGTGTCTGTG CCTTAATTCC AGCTATTATC TTGGCTGATG	2940
ATGTGGTGAC TTTGGATTGC GTTCCAGATA TTTCCGATGT AGCCAGTCTT GTCGAAATCA	3000
TGGAATTGAT GGGAGCTACT GTTAAGCGTT ATGACGATGT ATTGGAGATT GACCCAAGAG	3060
GTGTTCAAAA TATTCCAATC CCTTATGGTA AAATTAACAG TCTTCGTGCA TCTTACTATT	3120
TTTATGGGAG CCTCTTAGGC CGTTTTGCTG AAGCGACAGT TGGTCTACCG GGAGGATGTG	3180
ATCTTGTGCC TCGTCCGATT GACTTACACC TTAAGGCGTT TGAAGCTATG GGTGCCACTG	3240
CTAGCTACGA GGGAGATAAC ATGAAGTTAT CTGCTAAAGA TACAGGACTT CATGGTGCAA	3300
GTATTACAT GGATACGGTT AGTGTGGGAG CAACGATTAA TACGATGATT GCTGCGGTTA	3360
AAGCAATGG TCGTACTATT ATTGAAAATG CAGCCCCGTA ACCTGAGATT ATTGATGTAG	3420

185

CTACTCTCTT GAATAATATG GGTGCCCATATA TCCGTGGGGC AGGAACTAAT ATCATCATTATA	3480
TTGATGGTGT TGAAAGATTA CATGGGACAC GTCATCAGGT GATTCCAGAC CGCATTGAAG	3540
CTGGAACATA TATATCTTTA GCTGCTGCAG TTGGTAAAGG AATTCGTATA AATAATGTTC	3600
TTTACGAACA CCTGGAAGGG TTTATTGCTA AGTTGGAAGA AATGGGAGTG AGAATGACTG	3660
TATCTGAAGA CAGCATTTTT GTGAGGAAC AGTCTAATTT GAAAGCAATC AATATTAAGA	3720
CAGCTCCTTA CCCAGGCTTT GCAACTGATT TGCAACAACC GCTTACCCCT CTTTTACTAA	3780
GAGCGAATGG TCGTGGTACA ATTGTCGATA CGATTTACGA AAAACGTGTA AATCATGTTT	3840
TTGAACTAGC AAAGATGGAT GCGGATATTT CGACAACAAA TGGTCATATT TTGTACACGG	3900
GTGGACGTGA TTTACGTGGG GCCAGTGTTA AAGCGACCGA CTTAAGAGCT GGGGCTGCAC	3960
TAGTCATTGC TGGGCTTATG GCTGAAGGTA AAAGTGAAT TACCAATATC GAGTTTATCT	4020
TACGTGGTTA TTCTGATATT ATCGAAAAAT TACGTAATTT AGGAGCGGAT ATTAGACTTG	4080
TTGAGGATTA AACCGTAGAG GTGTTTATGA ATATTGGAC CAAATTAGCA ATGTTTCTCT	4140
TTTTTGAAAC GGATCGCTTG TATTTGCGTC CTTTCTTTT TAGTGATAGT CAGGACTTCC	4200
GCGAGATAGC TTCAAATCCA GAAAATCTTC AATTTATTTT CCCAACGCAG GCAAGTCTGG	4260
AAGAAAGTCA ATATGCACTG GCCAATTACT TTATGAAGTC CCCTTTGGGA GTGTGGGCAA	4320
TTTGTGACCA GAAAAATCAA CAAATGATTG GTTCTATTAA ATTTGAGAAG TTAGATGAAA	4380
TCAAAAAAGA AGCTGAGCTT GGCTATTTT TGAGAAAAGA TGCTTGGTGG CAGGATTTA	4440
TGACAGAGGT TGTTAGAAAA ATTTGTCAGC TTTCTTTTGA GGAATTTGGC TTAACAACAT	4500
TATTTATCAT TACCCACCTT GAAAATAAAG CTAGCCAAAG AGTTGCTCTT AAGTCTGGAT	4560
TTAGTTTGTT CCGTCAGTTT AAGGGAAGTG ATCGTTACAC AAGAAAAATG CGGGATTATC	4620
TTGAATTTTG GTATGTAATA GGAGAGTTCA ATGAGTAAGC ATCAGGAAAT TCTAAGCTAT	4680
TTGGAGGAAT TACCAGTAGG TAAAAGGGTC AGTCTTCGTA GCATTTTCGAA TCATCTAGGA	4740
GTTAGTGATG GAACAGCCTA TCGGGCTATT AAAGAAGCTG AAAACCGTGG AATTGTGGAG	4800
ACCCGTCCTA GAAGTGGAAC AATTCGTGTT AAATCCCAGA AAGTTGCTAT AGAGAGATTA	4860
ACGTTTGCTG AAATGTCAGA AGTGACTTCT TCTGAGGTTC TGGCTGGGCA AGAAGGTTTA	4920
GAGAGAGAA TTAGTAAGTT TTCAATTGGT GCCATGACTG AACAAAAAT CTTGTCTTAC	4980
CTTCATGATG GGGGGCTCTT GATTGTCGGA GACCGAACCC GTATTTCAGT GCTAGCCTTG	5040
GAAAATGAAA ATGCAGTTCT GGTACAGGG GGATTTCAGG TTCATGATGA TGTGCTTAAA	5100
CTGCCCAATC AAAAAGGGAT TCCTGTTCTA AGAAGTAAGC ATGATACCTT TACCGTCGCG	5160



186

ACCATGATCA ATAAAGCCTT GTCAAATGTC CAAATCAAGA CTGATATTCT GACAGTTGAG	5220
AAACTTTATC GCCCTAGTCA TGAGTATGGT TTTCTGAGAG AGACAGATAC AGTTAAAGAT	5280
TATTTGGACT TGGTTCGTAA GAATCGTAGC AGCCGTTTCC CTGTTATCAA TCAACATCAG	5340
GTCGTTGTTG GTGTTGTAAC CATGAGAGAC GCTGGTGATA AATCACCAAG CACGACAATT	5400
GATAAGGTTA TGTCTCGTAG TCTATTTTGT GTTGGATTAT CGACAAATAT TGCCAATGTG	5460
AGTCAACGGA TGATCCGAGA AGACTTTGAA ATGGTACCAG TTGTTCGAAG CAATCAAAC	5520
TTGCTTGGCG TTGTGACGCG ACGAGATGTC ATGGAGAAGA TGAGCCGTTT CCAAGTTTCC	5580
GCTCTACCAA CTTTTCTGA GCAGATTGGA CAAAAGCTCT CTTATCACCA TGATGAAGTA	5640
GTCATTACAG TGGAAACCTT TATGCTAGAA AAAAATGGAG TTTTGGCTAA TGGTGTATTG	5700
GCAGAAATTC TGACCCACAT GACCCGATT AGTTGTTAAT AGTGGTCGCA ATCTCATTAT	5760
CGAGCAGATG CTGATCTACT TTTTGCAGGC TGTTTCAGATA GATGATATAT TGCGCATTCA	5820
GGCACGGATT ATTCATCATA CGAGACGGTC AGCTATAATT GATTACGATA TTTATCATGG	5880
TCACCAGATT GTTTCAAAG CAAATGTGAC TGTTAAATTT AATTAGAAAC TAGGAGAAAA	5940
GATGATAACA TTAATATCAG CTCGTGAAAT CGAAGCTATG GACAAGGCTG GTGATTTTCT	6000
AGCAAGTATT CATATAGGCT TACGTGATTT GATTAAGCCA GGCCTAGATA TGTGGGAAGT	6060
TGAAGAATAT GTCCGCCGTC GTTGTAAGA AGAAATTTT CTTCCACTTC AGATTGGGGT	6120
TGACGGTGCC ATGATGGACT ATCCTTATGC TACCTGTTGC TCTCTTAACG ATGAAGTGGC	6180
TCACGCTTTC CCTCGTCATT ATATCTTGAA AGATGGTGAT TTGCTCAAAG TTGATATGTT	6240
TTTGGGAGGT CCCATTGCTA AATCTGACCT AAATGTCTCA AAATTAACT TCAACAATGT	6300
TGAACAAATG AAAAAATACA CTCAGAGCTA TTCTGGTGGT TTAGCAGACT CATGTTGGGC	6360
TTATGCTGTT GGTACACCGT CCGAAGAAGT CAAAACTTG ATGGATGTAA CCAAAGAAGC	6420
TATGTACAAG GGTATTGAGC AAGCTGTTGT TGGAAATCGT ATCGGTGATA TCGGTGCGGC	6480
TATTCAAGAA TACGCTGAAA GTCGTGGTTA CCGTGTAGTG CGTGATTGGG TTGGTCATGG	6540
TGTTGGCCCC ACTATGCACG AAGAACCAAT GGTTCCTAAC TATGGTATTG CAGGTCGTGG	6600
ACTCCGTCTT CGTGAAGGAA TGGTCTTAAC CATTGAACCA ATGATCAATA CAGGCGATTG	6660
GGAAATTGAT ACAGATATGA AAAGTGGTTG GGCGCATAAG ACCATTGACG GTGGATTGTC	6720
ATGTCAGTAT GAACACCAAT TTGTCATTAC GAAAGATGGA CCTGTTATCT TGAAGAGCCA	6780
AGGTGAAGAA GGAACCTATT AATAAAAAGT GAAAAGACTA CTGGAAGTTT ATTTTGATAA	6840
AAAATCCAGT AGATCTTTTC ATAATAAAC GCATTGTATC AAGTGTTAGG GGCTGATATC	6900
ATGCGTTTTT CTGCTTTTAA GATTTTCTCC AACTCTGTTT GTAAGCGCAT CATAACAAAG	6960

187

GGTCTAGGAT TCAGGGCTCT CCTCCTATAT ACTATTAGTA AAGTAAACT AAGGGAGGAT	7020
ATTTTAGTGT CGCAGTCTAT TGTTCCTGTA GAGATTCCAC AATATTGTCG TTTTGATTCT	7080
AAAAAGAGAA ATGGAATTCT GTTTAATGTT CGTATTGCCA ATCTTAAATT TACTTTTTTA	7140
TATTATACTT CCTGCGAAAC AAAATATGGT ATAGTAGTTC TATGAATGAT GAAGCAAGTA	7200
AACAACCTAAC TGATGCACGA TTTAAGCGTC TTGTTGGTGT TCAGCGTACC ACTTTTGAAG	7260
AGATGTTAGC TGTATTAAAA ACAGCTTATC AACTTAAACA CGCAAAAGGT GGACGAAAC	7320
CTAAATTAAG CCTAGAAGAC CTTCTTATGC CCACTCTTCA ATAGTGCGAG AATATCGAAC	7380
TTATGAAGAA ATTGCGGCTG ATTTTGGTAT TCACGAAAGC AACTTTATCC GTCGGAGCCA	7440
ATGGGTTGAA ATAACCTTTG TTCAAAGTGG TTTTACGGTT TCAAGAACTC CTCTCAGTTC	7500
TGAGGACACG GTAATGATTG ATGCGACGGA AGTAAAAATC AATCGCCCTA AAAAAACAAT	7560
TAGCGAATGA TTCTGGTAAA AAGAAATTTT ACGCTATGAA GGCTCAAGCG ATTGTCACAA	7620
GTCAAGGGAG AATTGTTTTCT TTGGATATCG CTGTGAACTA TAGTCATGAT ATGAAGTTGT	7680
TCAAAATGAG TCGTAGAAAT ATCGAACAAG CTGGTAAAAT CTTGGCTGAC AGTGGTTATC	7740
AAGGGCTCAT GAAGATATAT CCTCAAGCAC AAACCTCCAGC TAAATCCAGC AAACCTCAAGC	7800
CGCTAACAGC TGAAGATAAA GCCTATAACC ATGCGCTATC TAAGGAAAGA AGCAAGGTTG	7860
AGAACATCTT TGCCAAAGTA AAAACGTTTA AAATATTTTC AACAACCTAT CGAAATCATC	7920
GTAAACGCTT CGGATTACGA ATGAATTTGA GTGCTGGTAT TATCAATCAT GAACTAGGAT	7980
TCTAGTTTTG CAGGAAGTCT ATTGAGGTAT TGAGCTAGTT TATGAAAAAA TTGGGTGAAA	8040
AGTCGAGTGT TTTAGAAACC CACAGTGTAG TATTCTAGTT TCAATCCACT ATATTTTGCT	8100
ACTCCCCGTA AAGTTTCTAT TTCCCTGAT TTCTGATATA ATAGAAATAT TGACTTCAAG	8160
AGTAAGGAAG AGAAGATGAA CGCATTATTA AATGGAATGA ATGACCGTCA GGCTGAGGCG	8220
GTGCAAACGA CAGAAGGTCC CTTGCTAATC ATGGCAGGGG CTGGTTCTGG AAAGACTCGT	8280
GTTTTGACCC ACCGTATCGC TTATTTGATT GATGAAAAGC TGGTCAATCC TTGGAATATC	8340
TTGGCCATTA CCTTTACCAA CAAGGCTGCG CGTGAGATGA AAGAGCGTGC TTATAGCCTC	8400
AATCCAGCGA CTCAGGACTG TCTGATTGCG ACCTTCCACT CCATGTGTGT GCGTATTTTG	8460
CGTCGCGATG CGGACCATAT TGGCTACAAT CGTAATTTTA CAATTGTGGA TCCTGGTGAA	8520
CAGCGAACGC TCATGAAACG TATTCTCAA CAGTTGAACT TGGACCCTAA AAAATGGAAT	8580
GAACGAACTA TTTTGGGGAC CATTTCCAAT GCTAAGAATG ATTTGATTGA TGATGTTGCT	8640
TATGCTGCCC AAGCTGGCGA TATGTATACG CAAATTGTGG CCCAGTGTTA TACAGCCTAT	8700

188

CAAAAAGAAC TTCGTCAGTC TGAATCCGTT GACTTTGATG ATTTGATTAT GCTGACCTTG	8760
CGTCTCTTTG ATCAAAATCC TGATGTTTTG ACCTACTACC AGCAAAATTT CCAATACATC	8820
CACGTTGATG ACTACCAAGA TACCAACCAC GCTCAGTACC AATTGGTCAA ACTCTTGGCT	8880
TCCCGTTTTA AAAATATCTG TGTGGTTGGG GATGCGGACC AGTCTATCTA CGGTTGGCGT	8940
GGTGCTGATA TGCAGAATAT CTTGGACTTT GAAAAGGATT ACCCCAAAGC CAAGGTTGTT	9000
TTGTTGGAGG AAAATTACCG CTCAACCAAA ACCATTCTCC AAGCGGCCAA CGAGGTTATT	9060
AAAAATAATA AAAATCGCCG TCCTAAAAAT CTCTGGACTC AAAACGCTGA TGGGGAGCAA	9120
ATCGTTTACT ATCGTGCCGA TGATGAGCTG GATGAGGCTG TATTGTAGC CAGAACCATC	9180
GATGAACCTA GTCGCAGTCA AAACCTCCTT CATAAGGATT TTGCAGTTCT CTATCGGACT	9240
AATGCCCAGT CCCGTACAAT TGAGGAAGCC CTGCTCAAGT CTAACATTCC TTATACCATG	9300
GTTGGCGGAA CCAAATCTA CAGCCGTAAG GAAATTCGCG ATATTATTGC TTATCTCAAC	9360
CTTATTGCTA ATTTGAGTGA CAATATTAGT TTTGAGCGTA TTATCAACGA GCCTAAACGT	9420
GGAATTGGTC TAGGTACACT TGAGAAAAATC CGTGATTTTG CAAATTTGCA AAATATGTCT	9480
ATGCTGGATG CTTCTGCTAA TATTATGTTG TCTGGTATCA AGGGTAAGGC AGCCCAATCT	9540
ATCTGGGATT TTGCCAATAT GATGCTTGAT TTGCGGGAGC AGCTAGACCA CTTAAGCATT	9600
ACAGAGTTGG TTGAGTCCGT CCTAGAAAAA ACAGGTTATG TCGATATTCT TAACTCCCAA	9660
GCGACTCTAG AAAGCAAGGC ACGGGTTGAA AATATCGAAG AGTTTCTTTC TGTTACGAAG	9720
AACTTTGATG ACACCACGGA TGTGACAGAA GAGGAAACTG GTCTGGACAA ACTGAGTCGT	9780
TTCTTAAATG ACTTGGCTTT GATTGCCGAC ACAGATTGAG GTAGTCAGGA GACATCAGAA	9840
GTGACCTTGA TGACCTGCA TGCTGCCAAA GGTCTCGAAT TTCCAGTTGT CTTTTTGATT	9900
GGGATGGAAG AAAATGTCTT TCCACTTAGT CGTGCGACTG AAGATTGAGA TGAATTAGAA	9960
GAAGAGCGCC GTCTAGCCTA TGTAGGTATC ACGCGTGCAG AGAAAATTCT CTATCTGACC	10020
AATGCCAACT CACGCTTGCT TTTTGGTCGT ACCAATTATA ACCGTCCGAC TCGTTTTATT	10080
AACGAAATCA GTTCAGACTT GCTTGAGTAT CAAGGTCTGG CTCGTCCTGC AAATACAAGC	10140
TTTAAGGCAT CATATAGCAG TGGTAGTATT TCCTTTGGTC AAGGTATGAG TTTGGCTCAG	10200
GCTCTTCAAG ACCGTAAACG CGGTGCTGCC CCAAAATCAA TCCAGTCAAG CGGTCTTCCA	10260
TTTGGTCAAT TTACAGCTGG CGCAAAACCA GCATCTAGCG AGGCAAATTG GTCCATTGGT	10320
GATATTGCTC TCCACAAGAA ATGGGGAGAG GGAACCGTTC TCGAAGTTTC AGGTAGCGGT	10380
GCTAGGCAGG AATTGAAAAAT CAATTTCCCA GAAGTAGGTT TGAAAAAACT TTTAGCCAGT	10440
GTGGCTCCAA TTSAGAAAAA AATCTAATTT TCCATCCTTC TCACGAATAA TAAAGTGAGG	10500

AGGATTTTTA TGTACAGTAT TTCATTCCAA GAAGATTAC TATTACCAAG AGAAAGGCTG	10560
GCCAAGGAAG GAGTTGAAGC GCTTAGTAAC CAAGAGTTGC TAGCTATTTT ACTCAGGACA	10620
GGAACACGTC AAGCTAGCGT TTTTGAAATT GCCCAAAAAG TCTTGAACAA TCTTTCAAGC	10680
CTAACGGATT TGAAAAAAT GACCCTGCAG GAATTGCAGA GTTGTCTGG TATTGGGCGT	10740
GTAAAGGCCA TAGAATTACA AGCTATGATT GAACTGGGGC ATCGTATTCA CAAACACGAG	10800
ACTCTTGAAA TGGAAAGTAT TCTCAGCAGT CAAAGTTGG CCAAGAAGAT GCAGCAGGAA	10860
TTAGGGGATA AAAACAAGA GCACCTGGTG GCACTCTATC TCAATACTCA AAATCAAATC	10920
ATCCATCAGC AGACCATTTT TATCGGTCT GTAACCTGTA GTATCGCTGA ACCGCGAGAG	10980
ATTCTTCACT ATGCAATCAA GCATATGGCG ACTTCTCTTA TCTTGGTCCA CAATCATCCT	11040
TCAGGAGCGG TAGCGCCTAG CCAAAATGAT GATCATGTCA CTAAACTTGT TAAAGAAGCC	11100
TGCGAATTGA TGGGGATTGT TCTCTTGAC CATTGATTG TCTCTCATTC TAATTACTTT	11160
AGTTATCGTG AAAAGACAGA TTTAATCTAA AGTTCATTAA CGACATAGTC AAAGAGTTTT	11220
TTATCTTTGG GACGATTTTC AAAAGAAGT TCTGGATGCC ATTGGACACC GAGAAAGGCG	11280
ACATCATCCG TACTCATGAC AGCCTCAATG ATACCATCTT TAGGATCATG AGCCACAAC	11340
TTTAAATTTG GTGCTAAGTC CTTGATGCTC TGGTGGTGA AGGAGTTGAT ATGAGAGATT	11400
TCTCCATAGA TTTCTTGAG AACGGTATCT GGTCTGTGA CCAAGCGTTG AGTTGTGTAC	11460
TCAACAGAAG AATCCTGCCA ATGGTCTTCG ATATCTTGGT ACAAAGTTCC ACCCATGGCA	11520
ACGTTAAAGA GTTGGGTACC ACGGCAGACA GAGAAAATGG GCTTTTTCTG TTTAATAGCT	11580
TCCTTGATGA GGGCCAGTTC GAAGATATCT CTTTGAAGGT GATAGTCATC ACTATCAATG	11640
GTTTTGGGTT CGCCATAAAA TTTTGGATCG ACATTTTGCC CACCTGTCAA GATGAGCTTG	11700
TCAATCAAAC TGATATAGTG GCAGGCCATT TCTTGATCAC CAATCGGTAG GATGATGGGA	11760
ATCCCTCCAG CATCTTTAAC GCCTTCAACA AAGCCTTTTG CTGCGTAGCT CATCATGATG	11820
TCATCATCTG GATGAGTTTT TTCGTTTCCT GTAATCCCAA TAACTGGTTT TTTCATAAAA	11880
TGATTTTCGC TTTCTAATCC TCTTTTCGCA TGAAGTAGAG GAGGGTTTGG AGTTCACTTG	11940
TCAAATCGAC ATACTGAACG ACCACGTCTT TTGGTAAATG CAGATGGACT GGTGAAAAAC	12000
TGAGAATTCC TTTACACCA GCATCAACCA AGAGATTAGC AACCTCTTGT GACTTGACGC	12060
TGGGAACAGT TAGGATAGCA GTCTTCACAT CAGCATCCTT GATTTTATCC TTGATCTGAG	12120
AAATCCCGTA AATGGGAATC CCGTCAGGAG TTTGGGTACC GACTTCAGGA TGGTCGTCTA	12180
GGTCAAAGGC CATGATAATC TTCATCTTGT TACGTTCTGT GAAGCGGTAG TGGAGAAGGG	12240

190

CATGGCCCAT	ATTTCCAATA	CCAACCAGCA	TGACATTGGT	AATAGAGTTG	TCATTGAGCA	12300
AATCGGCAAA	AAATGTCATT	AGTTMTTGA	CATCATAGCC	AAAACCACGA	CGACCAAGTT	12360
CACCAAAATA	GGAAAAATCA	CGACGTACGG	TCGCTGAATC	AATACCGATA	GCCTCTGCAA	12420
TTTGCTTAGA	GTTGGCACGT	TCAATCTTTT	CTGCATGAAA	TCTCTTAAAA	ATTCGATAGT	12480
AGAGAGAGAG	TCTTTTGTCT	GTAGCTTTTG	GAATAGCAAA	CTGTTTATCT	TTACACAAAT	12540
CACAACCTTT	CTATTCTTCT	ATTTTATAGA	AACATTGTGA	AAAAATCAAC	AAAAATAAGA	12600
AAAAACTAAG	AAAAATCTTA	GTTTTGATGT	AAAAATCTG	CATGAGATAG	AAAACGGTAG	12660
AGGTCTCCGA	CCAGCCCCTG	ATAAACTTTT	TTGCCCCATA	AAGTCAGAGA	AGTCACATAA	12720
ACTGTATCTG	GTAAGGTAC	ACATCCTGAC	AAAGTCAACA	TGAGAGCCTC	ATGATCCTCA	12780
TACTTGAGAG	TACGCTCTAC	ATGATAGCAG	TCCTTATAGG	TCAGTTCAAA	CATTTTGGCT	12840
CTATCTTTCC	GATTTTGTA	AGACACCACG	TTCTACCAAG	CTATCCATGA	GGAAGTAGAA	12900
TTTTTCCTGA	TGAATATGGT	GGTCTTCTGA	TTTGAAAATA	TCAACTAGAC	GAAGGCCAAA	12960
CTTGTCAGTG	ATATTGATTT	TAGCCCCTGT	AAGTTCCTTG	TTAATGATGA	TTTTGAGTTG	13020
GAAGCCTTCA	CCGCTGTTTG	GCACTTTTTC	CAAAAGGCCA	GTCAATTTCAT	AGTTACCAAC	13080
CTTAGTTTCA	AAAAAGGTGT	TATCTTTGAG	GGTGAATTTT	TTAACAGAAG	GGCTAAGACT	13140
GTAATCGTAA	CGACAATTTT	TTAACTGAAT	GATTTTTTCA	AATGCCATAT	GGCTAACCTC	13200
CGATAATTTT	TTTTAAGGTT	TTTGCGAGGG	TTTGTAGGTC	TTCAACGGTA	TTTTGTGGCG	13260
ACAAACTGAT	GCGAAGGGAT	TCCTTCAAGC	GTTCTGAATT	TGCGCCATAC	ATGGCTTCAA	13320
GAACATGGCT	GGATTGGACA	ACGCCTGCAG	TACAGGCTGA	GCCAGTAGAG	ATTGAAATTC	13380
CAGCTAAATC	TAGCCGAAGG	AGTAAGAGGT	CATTTTTCTG	ACCAGGAAAT	CCAATATTGA	13440
GAACATAAGG	GAGATGATGT	TTTCCTCTAT	TCAGGTAATA	CTGAATGCCC	TCCAGCTCTG	13500
CCAGAAAGGC	AGTTTCTAGA	TTTTGTACAT	GTTGAAAATG	TTCTTCTTGT	TTTTCTAGGT	13560
CTTCTTTTAG	GGCTGCAACC	ATGCCTACAA	TGGCAGGCAG	ATTTTCAGTT	CCTGCACGTT	13620
TTTTCTGTTT	CTGGTCTCCG	CCATGTAGAT	AGGAATCAAA	GTCCATGCTA	GATGCGTAGA	13680
GAAAACCGAT	TCCCTTAGGA	CCATGGAATT	TGTGGGCAGA	AGCAGTGAGA	AAATCAATGC	13740
CCAATTCCTC	TGAATGAATT	GGGATTTTAC	CAATAGCCTG	AACTGCATCA	ACATGATAGG	13800
CAGCAGGGTG	TTGCTTGAGT	ATTTGGCCAA	TTTCAGCGAT	GGGCAGTAGG	TTTCCTGTCT	13860
CATTATTGAC	AAACATGGTA	GAAACCAAAA	TCGTATCGTC	ACGTAAAGCC	TTTTGAATTT	13920
GCTGGGCTGT	GATTTCTTGA	TTTTCTGGCT	GGATAATGGT	TGCTTCAAAC	CCAAAGTGTT	13980
GAACCAAGTA	ATCAATTGTT	TCAAGGACAG	CATGCTGCTC	GATGGCAGTT	GTGATGATAT	14040

GTTTTCTTGG	TTCTTGGTGA	CGAAGACAGT	AGCCAATGAT	GGTAGTATTA	TTGCCTTCAG	14100
TCCCACCAGA	AGTGAAAAAG	ATATGTTGAG	GTTTGTCTCT	TAGTAACTGG	GCTAGTTCCT	14160
GACGGGCTTC	TCGCAAGAGT	TTGCCAGCTT	GACGACCATG	ACCATGAATA	CTAGAAGGAT	14220
TTCCGTGGGT	TTCTTGCAATA	ACCTTGGTCA	TAGCTGAAAT	AGCAACTGCT	GACATAGGAG	14280
TCGTTGCAGC	ATTGTCCAAA	TAAATCAAAG	AATCACCTTA	TTTCTTTTTA	TTGTAGGCAA	14340
AGAGTGGGCT	GACTGGTTTT	CTTTCGTGAA	TACGGACGAT	AGCATCACCA	ATTAACCTCAC	14400
TAGCAGTGAT	GTAGCATACA	TTTTTAGGAG	TTTTTCTTTT	TGTTGCTACT	GAATCAGTCA	14460
CAAGAATTTC	TTTAATATTA	GTATTGTCAA	GAAGTCAGC	AGCTCCCTCG	ACGAAGAGAC	14520
CGTGGCTAGA	AACAGCATAA	ATTTCTGTAG	CTCCTTCACG	TTCAACGATT	TTAGAAGCTT	14580
CAGAGAAGGT	ACGTCCTGTA	TTTAAATAT	CATCAATCAA	GATAGCTTTC	TTACCTTCAA	14640
CATCACCAAT	AATATAACCT	TCGTTACGAG	TTGCATCGTC	TTGAGGGTAG	TCGATAATGG	14700
CGATAGGAGC	ATCAAGATAT	TCAGCCAGGC	TACGGCGACG	TTTGACACCT	GAATTTTTAG	14760
GGCTAACGAC	AACAACATCT	GAACCAAGCA	ATCCTTTATC	GCAGTAATGT	TTTGCGAATA	14820
GGGGAACAGT	GAAAAGATTA	TCCACTGGAA	TATCAAAGAA	ACCTTGAACC	TGAACGGCAT	14880
GCAATCAAG	AGTCAGGATA	CGATCAACTC	CAGCCTTAAC	CAGCATATTG	GCAACTAGTT	14940
TTGCTGTAAG	TGGCTCACGA	GGACAAGCAA	TGCGGTCTTG	ACGTGCATAG	CCAAAATATG	15000
GAAGGACAAC	GTTGATACTG	TGGGCACTTG	CACGCACACA	AGCATCGACC	ATGATTAACA	15060
ATTCCATTAG	GTGGTTGTTG	ACAGGGAAAC	TTGTTGATTG	GATGATGTAA	ACATCATAAC	15120
CACGGACACT	TTCTTEGATA	TTTACTTGGA	TTTCTCCGTC	TGAAAATTGA	CGTGATGATA	15180
GTTTTCCAAG	TGGGACACCA	ACAGCTTGGG	CAATTTTTTG	TGCAATCTCT	TCGTTAGAGT	15240
TGAGTGGCAA	AAGTTTCATG	TTTTTCTAT	CTGACATTAT	AGACCGTCCT	CTGTAAACTT	15300
TATAAATCCT	AGTTATATTT	ACCTTACATA	TATGAACTGG	GATTTGTGTA	TTTTTATCTT	15360
TTCTATTTTA	CCAAAAATG	GAGATTATTT	CAGCTATTTT	TCATACTTTT	GACAAATCGA	15420
ACCAATTTTG	AAGGAGCTTT	TTGATAGGAA	ATCTGATTTT	TCTCTAAAAA	TTGTGCAAAA	15480
TCCTGTTTGC	CTTGCTCATG	ATTTTCCACT	TCAAGCTCCA	ATTCGTAATC	TGTTATATCA	15540
AAGTATCGGC	TCTGATCCAG	TGCCATGAGA	CCAATAGCTG	TTTTTCATTT	ATAGCGAAGC	15600
GTGTGTTAGAC	AACCAAGAAC	CTGCCAGTTC	TACTTTTGGA	TACCATGTTT	CGCCAATTCA	15660
TCCAGTACTA	GGCCTTGAGG	AAGTTCTTCC	TACTCAGAT	AGTTCTCAGC	ATCTTTTAGT	15720
TGCAATTTT	GGTTGTATTC	CATGTTTCCA	ACACTCTGCG	GGACTTTGAG	TGTCAACTCA	15780

192

GCCCAGTCTT CAAAGGTTCC AATGCCGATA GCGACTTTCT TTTCTC CAG TTCAAAATCA	15940
GCGGTGTCTGA TGTAGTAATT TGTTTGAAGA ACAGGAGTGA CACCTGTGAA CTGGTCTTTT	15900
AGACGATTGT ATTCATCTTT TTTCAATAGT GTTTTCAATT CAATTTCTAA ATGTTTCATT	15960
TTTCTTACCT TTTTTTATCG TTGAAAGCGG ATTTATGGTA TAATAAGCAT TGTATTTATT	16020
GTATATGAAT CTGGAGAAAA AATCAAAGAT ATTTTGTACG GATAATATGA GAACAAGGGA	16080
GAATATATGA CCTTAGAATG GGAAGAATTT CTAGATCCTT ACATTCAAGC TGTGCTGAG	16140
TTAAAGATTA AACTTCGTGG TATTCGTAAG CAATATCGTA AGCAAAATAA GCATTCTCCA	16200
ATTGAGTTTG TGACCGGTGC AGTCAAGCCA ATTGAGAGCA TCAAAGAAAA AATGGCTCGT	16260
CGTGGCATT AATTATGCGAC CTTGGAACAC GATTTGCAGG ATATTGCTGG CTTACGTGTG	16320
ATGGTTTCAGT TTGTAGATGA CGTCAAGGAA GTAGTGGATA TTTTGCACAA GCGTCAGGAT	16380
ATGCGAATCA TACAGGAGCG AGATTACATT ACTCATAGAA AAGCATCAGG CTATCGTTCC	16440
TATCATGTGG TAGTAGAATA TACGGTTGAT ACCATCAATG GAGCTAAGAC TATTTTGGCA	16500
GAAATTCAAA TTCGTACTTT GGCCATGAAT TTCTGGGCAA CGATAGAACA TTCTCTCAAC	16560
TACAAGTACC AAGGGGATTT CCCAGATGAG ATTAAGAAGC GACTGGAAAT TACAGCTAGA	16620
ATCGCCCATC ACTTGGATGA AGAAATGGGT GAAATTCGTG ATGATATCCA AGAAGCCAG	16680
GCACTTTTTG ATCCTTTGAG TAGAAAATTA AATGACGGTG TAGGAAACAG TGACGATACA	16740
GATGAAGAAT ACAGGTAAAC GAATTGATCT GATAGCCAAT AGAAAACCGC AGAGTCAAAG	16800
GGTTTTGTAT GAAATGCGAG ATCGTTTGAA GAGAAATCAG TTTATACTCA ATGATACCAA	16860
TCCGGATATT GTCATTTCCA TTGGCGGGGA TGGTATGCTC TTGTCGGCCT TTCATAAGTA	16920
CGAAAATCAG CTTGACAAGG TCCGCTTTAT CGGTCTTCAT ACTGGACATT TGGGCTTCTA	16980
TACAGATTAT CGTGATTTTG AGTTGGACAA GCTAGTGAAT AATTTGCAGC TAGATACTGG	17040
GGCAAGGGTT TCTTACCCTG TTCTGAATGT GAAGGTCTTT CTTGAAAATG GTGAAGTTAA	17100
GATTTTCAGA GCACTCAACG AAGCCAGCAT CCGCAGGTCT GATCGAACCA TGGTGGCAGA	17160
TATTGTAATA AATGGTGTTT CCTTTGAACG TTTTCGTGGA GACGGGCTAA CAGTTTCGAC	17220
ACCGACTGGT AGTACTGCCT ATAACAAGTC TCTTGGCGGT GCTGTTTTAC ACCCTACCAT	17280
TGAAGCTTTG CAATTAACGG AAATTGCCAG CCTTAATAAT CGTGTCTATC GAACACTGGG	17340
CTCTTCCATT ATTGTGCCTA AGAAGGATAA GATTGAACTT ATTCCAACAA GAAACGATTA	17400
TCATACTATT TCGGTTGACA ATAGCGTTTA TTCTTTCCGT AATATTGAGC GTATTGAGTA	17460
TCAAATCGAC CATCATAAGA TTCACTTTGT CCGGACTCCT AGCCATACCA GTTTCTGGAA	17520
CCGTGTTAAG GACGCCTTTA TCGGCGAGGT GGATGAATGA GGTTTGAATT TATCGCAGAT	17580

GAACATGTCA AGGTAAAGAC CTTCTTAAAA AAGCACCAGG TTTCTAAGGG ATTGCTGGCC	17640
AAGATTAAGT TTCGAGGTGG AGCTATTCTG GTCAATAATC AACCGCAAAA TGCAACGTAT	17700
CTATTGGACG TTGGAGACTA CGTTACCAT TACATTCCCG CTGAGAAAGG CTTTGAAACC	17760
TTGGAGGCTA TTGAGCTTCC ATTAGATATT CTCTATGAGG ATGACCACTT TCTAGTCTTG	17820
AATAAACCCCT ATGGAGTGGC TTCTATTCCCT AGTGTCAATC ACTCTAATAC CATTGCCAAT	17880
TTTATCAAGG GTTACTATGT CAAGCAAAAT TATGAAAATC AGCAGGTTCA CATTGTTACC	17940
AGACTAGATA GGGATACTTC TGGCTTGATG CTCTTTGCCA AGCACGGTTA TGCCCATGCA	18000
CGATTAGACA AGCAGTTGCA GAAGAAATCT ATCGAGAAAC GCTACTTTGC TTTGGTTAAG	18060
GGAGATGGAC ATTTGGAGCC AGAAGGGGAA ATTATTGCTC CGATTGCCGC TGATGAAGAT	18120
TCCATTATTA CCAGACGAGT GGCTAAAGGC GGAAAGTATG CCCATACTTC ATACAAGATT	18180
GTAGCTTCTT ATGGAAATAT TCACCTGGTC TATATTACC TGCACACTGG TCGAACCCAT	18240
CAAAATCCGAG TCCATTTTTC TCATATCGGT TTTCTTTGC TGGGAGATGA TTTGTATGGT	18300
GGTAGTCTGG AAGATGGTAT TCAACGTCAG GCTCTGCATT GCCATTACCT ATCCTTTTAT	18360
CATCCATTTT TAGAGCAAGA CTTGCAGTTA GAAAGTCCCT TGCCGGATGA TTTTAGTAAC	18420
CTTATTACCC AGTTATCAAC TAATACTCTA TAAAACTGT CTCAGAGTAT AATTATTATC	18480
TTAAAGGAGA AAATCATGG AAGTTTTTGA AAGTCTCAA GCCAACCTTG TTGGTAAAAA	18540
TGCTCGTATC GTTCTCCCTG AAGGGGAAGA GCCTCGTATT CTTCAGCAA CAAAACGCTT	18600
AGTAAAAGAA ACAGAAGTGA TTCCTGTTTT GCTTGGAAAT CCTGAAAAAA TTAATAATTA	18660
TCTTGAAATT GAAGGAATCA TGGATGGTTA TGAGGTCATC GACCCCTAAC ATTATCCTCA	18720
ATTTGAAGAA ATGGTTTCTG CCTTGGTGGG GCCTCGCAAG GGCAAAATGA CTGAAGAAGA	18780
TGTACGCAAG GTTTTGGTTG AAGATGTCAA CTACTTTGGT GTGATGTTGG TTTACTTGGG	18840
CTTGGTTGAT GGAATGGTGT CAGGAGCGAT TCACTCAACA GCTTCAACAG TTCGCCCAGC	18900
TCTACAAATC ATCAAAACTC GTCCAAATGT AACTCGTACT TCAGGAGCCT TCCTCATGGT	18960
TCGTGGTACG GAACGTTACC TATTTGGAGA CTGTGCCATT AACATCAATC CAGATGCAGA	19020
AGCCTTGGCT GAAATTGCCA TCAACTCAGC AATCACAGCT AAGATGTTTG GCATCGAACC	19080
TAAATTTGCC ATGTTGAGCT ATTCTACTAA AGGTTCAAGG TTTGGTGAAA GCGTTGATAA	19140
GGTCGTTGAA GCAACTAAAA TTGCTCAGCA CTTGCGTCCT GACCTTGAAA TCGATGGTGA	19200
GTTGCAATTT GATGCAGCCT TTGTTCTGA AACTGCAGCT CTGAAAGCTC CTGGAAGTAC	19260
GGTAGCTGGT CAAGCAAATG TCTTCATCTT CCCAGGTATC GAGGCAGGAA ATATTGGTTA	19320



194

CAAGATGGCT GAACGCCTGG GTGGCTTTGC GGCTGTAGGA CCTGTTTTGC AAGGTTTAAA	19380
CAAGCCAGTT AATGATCTTT CTCGTGGATG TAATGCAGAT GATGTTTACA AGTTGACCCT	19440
CATCACAGCA GCTCAAGCAG TTCATCAATA GTGAAACTA TAAAGTGATA TACTATGCTA	19500
TACTGTAGTT ATGAACTAT GTACGAAAAG CACTGCCATT AATTCCTGAG AACTAAATTA	19560
CTGATTGGTG TCAAAAAGGA AAACCTCCAA GCGATGATAT CCTGTCTATA CACGACCTAT	19620
AGAAATCTGT AATATACATA TCCGTAAAAC GATAAATTCC CTTTTTGATT TTAATGAGT	19680
ATGAAAAGAG AATTTTTTGG CTCTTTGTCA ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC	19740
GAGAAAGGAC AAATTTTCATC CTTCTTTTTT TGATATTCAG AGCGATAAAA ATCCGTTTTT	19800
TGAAGTTTTT AAAGTTCCGA AAACCAAAGG CATTGCGCTT GATAAGTTTG ATGAGATTAT	19860
TGGTGGCTTC CAGTTTGGCG TTAGAATAGT GTAGTTGAAG GCGTTTGATA ATCTTTTCTT	19920
TATCTTTGAG GAAGGTTTTA AAGACAGTCT GAAAAATAGG ATGAACCTGC TTAAGATTGT	19980
CCTCAATAAG TCCGAAAAAT TTCTCTGGTT CCTTATCTCG GAAGTGAAAA AGCAAGAGTT	20040
GATAGAGCTG ATAGTGGTGT TTCAAGTCTT CCGAATAGCT CAAAAGCTTG TTTAAATCT	20100
CTTTATTGGT TAAGTGCATA CGAAAAATAG GACGATAAAA TCGCTTATCA CTCAGTTTAC	20160
GGCTATCCTG TTGAATGAGT TTCCAGTAGC GCTTGATAG	20199

## (2) INFORMATION FOR SEQ ID NO: 7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19702 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ACCCGATGTA TCAGCGGATA TTTACTCTAT TTTTCAAACG ATGTTATACC CACAATAAAA	60
GAAAAAAGAC CCTAAGGTCT CCTTTGCTTT TATTATTAAA CGCGTTCAAC TTTACCTGAT	120
TTCAAAGCAC GAGCTGAAGC CAAAACTTTT TTAGGTTTAC CATCGATAAG AACAGTAACT	180
TTTTGAAGGT TTGGTTTTAC GGCACGTTTT GTTTGGTTCA TCGCGTGTGA ACGGTTGTTT	240
CCTGATACAG TCTTACGACC TGTAAGTAA CATACTTTAG CCATTGTGTT TTCCTCCTAT	300
TAGATCTAAT ATAGCGGATG TGCTAGCACC ACATACCGTA CTATGTTATC ACATTTTCTT	360
GTTTTTTCGA AGGSAATTGG AAGATTTTTT ATTTGTGTCT TAAATCAGGT CTTCGGTGAC	420
ATTTCTGCTC TCCACATGCC ATCGTTGATT AACAGAACAC CAGAATTAAT ATTATGTGTA	480
TAAAAATCAT CTCTAACTGC AGCTAAGGGT ATAGCCGTCA AGTCCAAATC CCACAGCTCA	540

195

TCTATCGATT TTCTTACAAC AATATCTGAA TCCAAATACA GTACACGAGA CTCGCTTACA	600
TACTTTGGAA TAAAATACCT AAAAAAGCCG CATATGAAAG TCCCTCAAAG GGGAGACGAT	660
AACCTTTCAG AATATTACTG TCAATCTAAA CATTACAAAT CTCACTATTC AAAGTCTCTA	720
GTCTTTTTC CATCAATTGG AACCATTCTC GCGGAAGGTC ATCATTAAAA ACATAAACT	780
TAAGATTATA ATGATGAACA CAAAGAGATT TTATTGTTGT TTCAACTTTA TCCATATAAG	840
CATTATCTGC ACCTAAGACA ATCGCTTTT TCTCTCTTT CACTTTTTAT CTCATTTCTT	900
TTTATTCCCA TCATATTATT CCCATCATAT GTTCCCATC ATATGTTTCT ACGTAACCAT	960
TATTTTCGCC TATTCGTTTCG TAAAACCATA CCAGTGGAGA TTTTAGATGA AGTCCCATTA	1020
CGGTTTACAA TTTTACATT ACGACACGGA GTTTTACAAA TCGATTTCAT TTGCCAAACG	1080
TAGTTAGTGA GGCAGTTAGC TAGTTCGCCA AATAGCGACT AGCGTCCAAC AATTTGGAAC	1140
TTTAGTTCCA ATTGTTGGTA CTGAGTCACA TCTTCTCTC TAACTCTACG TCTGGATACT	1200
TGTCCGCAAA CCAGCGGAGG GCAAAGTCAT TTTCAAAGAG AAAGACTGGT TGGTCAAAAC	1260
GGTCTTTGGC TAAGATATTG CGACTTGACG ACATCCGTTT ATCCAAGTCC TCAGGCTTGA	1320
TCCAACGAAC GGTCTTTTTA CCCATTGGGT TCATAACTAC TTCCGCATTG TACTCGCCTT	1380
CCATGCGGTG TTAAAGACT TCAAACCTGA GTTGACCTAC AGCGCCTAGC ATGTACTCAC	1440
CTGTTTGGTA ATTCTTATAA AGCTGAACGG CTCCTTCTTG CACCAATTGC TCAATCCCCT	1500
TGTGGAAGGA TTTTGCTTC ATAACATTCT TAGCAGAAAC TTTCATGAAA ATCTCAGGTG	1560
TAAAGGTTGG CAGGGGTCA AATTCAAAC TGTTTTTTCC AACCGTCAAG GTATCCCCAA	1620
CCTGATAAGT ACCGGTATCG TAAACCCCGA TAATATCACC TGCCACGGCA TTGGTCACAT	1680
TCTCAGCACT CTCCGCCATA AACTGGGTAA CATTAGATAG TTTAGGCTTC TTACAGTAC	1740
GAGGGAGATT GACACTCATG CCGCGCTCAA ATTCCGCCAGA TACGATACGG ACAAAGGCAA	1800
TACGGTCACG GTGACGAGGG TCCATGTTGG CTTGGATTTT AAAGACAAAG CCTGAGAAAT	1860
CCTTGTGATA AGGATCCACA ATTTACCCGT CTGTTTTCTT GTGACCATGT GGTCTGGAG	1920
CAAACCTGAG GAAGGTTTCA AGGAAGGTCT GCACACCAAA GTTTGTCAGG GCTGAACCGA	1980
AAAAGACAGG CGTCAATTCT CCAGCCAGAA TAGCTTCCTC TGAAAACCTA TTCCCGCTT	2040
CATTTAAAAG CTCAATGTCA TCCTTGACTT GCTCGTAGAA AGGATTGCTA CCAAAGAGTT	2100
TGTCCCCGTC TTCTAGACTG GCAAAACGCT CATCCCCCTT GTAAAGCTCT AAACGTTGGT	2160
TATAGAGGTC ATACAAGCCC TCAAAGGCTT TCCCCATCCC GATAGGCCAG TTCATAGGGT	2220
AGCTAGCAAT GCCCAAGATT TCTTCCAATT CTTGCAAGAG ATCCAAAGGC TCACGACCGT	2280

196

CACGGTCCAG CTTGTTTCATA AAGGTAAAGA CTGGAATGCC ACGATGTTTC ACAACCTCAA	2340
ACAATTTCTT GGTITGAGCC TCGATCCCCT TGGCAGAGTC CACGACCATG ACCGCAGCAT	2400
CCACCGCCAT CAAGGTACGA TAGGTATCTT CTGAGAAGTC CTCGTGCCCT GCGGTGTCTA	2460
AGATATTAC GCGCTTGCCG TCGTAGTCAA ATTGCATAAC AGATGAAGTA ACAGAAATCC	2520
CACGTTGCTT CTCGATATCC ATCCAGTCAG ATTTAGCAAA AGTCCCTGTT TTCTTCCCTT	2580
TTACCGTACC AGCCTCACGA ATCTCACCCC CAAAGTAGAG TAACTGCTCA GTGATGGTTG	2640
TTTTCCCCC GTCCGGGTGG GAGATAATGG CAAAGGTACG ACGTTTCTTA ATTTCTTCTT	2700
GAATATTCAT AAGTTCTCTT TCTTTGATC TCTATTTTTC TTGTTTCAAT AGCTGAGAAT	2760
GATTTTACA TTGGATTTTA CCATTCCTTT CAACACTCCA TTATATCGGA TTTTAGCATT	2820
TTTTTCAATT TCTATTTCTT TTCACTTCCC CCTCCCTTAT TTATAGGAAA ATATGGTAAA	2880
ATAGAACAGA CTAATAATCA TCATTTACAG AAAGGATGCA AGATGAAAAT TACGCAAGAA	2940
GAGGTAACAC ACGTTGCCAA TCTTTCAAAA TTAAGATTCT CTGAAGAAGA AACTGCTGCC	3000
TTTGCGACCA CCTTGTCTAA GATTGTTGAC ATGGTTGAAT TGCTGGGCGA AGTTGACACA	3060
ACTGGTGTG CACCTACTAC GACTATGGCT GACCGCAAGA CTGTACTCCG CCCTGATGTG	3120
GCCGAAGAAG GAATAGACCG TGATCGCTTG TTTAAAAACG TACCTGAAAA AGACAACCTAC	3180
TATATCAAGG TGCCAGCTAT CCTAGACAAT GGAGGAGATG CCTAATGACT TTTAACAATA	3240
AAACTATTGA AGAGTTGCAC AATCTCCTTG TCTCTAAGGA AATTTCTGCA ACAGAATTGA	3300
CCCAAGCAAC ACTTGAAAAT ATCAAGTCTC GTGAGGAAGC CCTCAATTCA TTTGTACCA	3360
TCGCTGAGGA GCAAGCTCTT GTTCAAGCTA AAGCCATTGA TGAAGCTGGA ATTGATGCTG	3420
ACAATGTCTT TTCAGGAATT CCACTTGCTG TTAAGGATAA CATCTCTACA GACGGTATTC	3480
TCACAACCTG TGCCTCAAAA ATGCTCTACA ACTATGAGCC AATCTTTGAT GCGACAGCTG	3540
TTGCCAATGC AAAAACCAAG GGCATGATTG TCGTTGGAAA GACCAACATG GACGAATTTG	3600
CTATGGGTGG TTCAGGTGAA ACTTCACACT ACGGAGCAAC TAAAAACGCT TGGGAACCACA	3660
GCAAGGTTC TGGTGGGTCA TCAAGTGGTT CTGCCGAGC TGTAAGCTCA GGACAAGTTC	3720
GCTTGTCACT TGGTTCTGAT ACTGGTGGTT CCATCCGCCA ACCTGCTGCC TTCAACGGAA	3780
TCGTTGGTCT CAAACCAACC TACGGAACAG TTTCACGTTT CCGTCTCATT GCCTTTGGTA	3840
GCTCATTAGA CCAGATTGGA CCTTTTGCTC CTAAGTTTAA GGAAAATGCC CTCTTGCTCA	3900
ACGCTATTGC CAGCGAAGAT GCTAAAGACT CTAAGTTCTG TCCTGTCCGC ATCGCCGACT	3960
TTACTTCAAA AATCGGCCAA GACATCAAGG GTATGAAAAT CGCTTTGCCT AAGGAATACC	4020
TAGGCGAAGG AATTGATCCA GAGGTTAAGG AAACAATCTT AAACGCGGCC AAACACTTTG	4080

AAAAATTGGG TGCTATCGTC GAAGAAGTCA GCCTTCCTCA CTCTAAATAC GGTGTTGCCG	4140
TTTATTACAT CATCGCTTCA TCAGAAGCTT CATCAAACCTT GCAACGCTTC GACGGTATCC	4200
GTTACGGCTA TCGCGCAGAA GATGCAACCA ACCTTGATGA AATCTATGTA AACAGCCGAA	4260
GCCAAGGTTT TGGTGAAGAG GTAAAACGTC GTATCATGCT GGGTACTTTC AGTCTTTCAT	4320
CAGGTTACTA TGATGCCTAC TACAAAAAGG CTGGTCAAGT CCGTACCCTC ATCATTCAAG	4380
ATTTCGAAAA AGTCTTCGCG GATTACGATT TGATTTTGGG TCCAACGCT CCAAGTGTG	4440
CCTATGACTT GGATTCTCTC AACCATGACC CAGTTGCCAT GTACTTAGCC GACCTATTGA	4500
CCATACCTGT AAACCTGGCA GGACTGCCTG GAATTTGAT TCCTGCTGGA TTCTCTCAAG	4560
GTCTACCTGT CGGACTCCAA TTGATTGGTC CCAAGTACTC TGAGGAAACC ATTTACCAAG	4620
CTGCTGCTGC TTTTGAAGCA ACAACAGACT ACCACAAACA ACAACCCGTG ATTTTGGAG	4680
GTGACAACTA ATGAACTTTG AAACAGTCAT CGGACTTGAA GTCCACGTAG AGCTCAACAC	4740
CAATTCAAAA ATCTTCTCAC CTACTTCTGC CCACTTTGGA AATGACCAAA ATGCCAACAC	4800
TAACGTGATT GACTGGTCTT TCCCAGGAGT TCTACCAGTT CTCAATAAAG GGGTTGTTGA	4860
TGCCGGTATC AAGGCTGCTC TTGCCCTCAA CATGGACATC CACAAAAAGA TGCACTTTGA	4920
CCGCAAGAAC TACTTCTATC CTGATAACCC CAAAGCCTAC CAAATTTCTC AGTTTGATGA	4980
ACCAATCGGA TATAATGGCT GGATTGAAGT CAAACTAGAA GACGGTACGA CCAAGAAAAAT	5040
CGGTATCGAA CGTGCCCAACC TAGAGGAAGA CGCTGGTAAA AACACCCATG GTACAGATGG	5100
CTACTCTTAT GTTGACCTCA ACCGCCAAGG GGTTCCTTGG ATTGAGATTG TATCTGAGGC	5160
AGATATGCGT TCTCCTGAAG AAGCCTATGC TTATCTGACA GCCCTCAAGG AAGTTATCCA	5220
GTACGCTGGC ATTTCTGACC TTAAGATGGA GGAAGGTTCC ATGCGTGTGG ATGCCAACAT	5280
CTCCCTTCGT CCTTATGGTC AAGAGAAATT CGGTACCAAG ACTGAATTGA AGAACCTCAA	5340
CTCCTTCTCA AACGTTCTGA AAGGTCTTGA ATACGAAGTC CAACGCCAGG CTGAAATTCT	5400
TCGCTCAGGT GGTCAAATCC GCCAAGAAAC ACGCCGTTAC GATGAAGCGA ATAAAGCAAC	5460
CATCCTCATG CGTCTCAAGG AAGGGGCTGC TGACTACCGC TACTTCCCAG AACCAGACCT	5520
ACCCCTCTTT GAAATTTCTG ACGAGTGGAT TGAGGAAATG CGGACTGAGT TGCCAGAGTT	5580
TCCAAAAGAA CGTCGTGCGC GTTATGTATC TGACCTTGGT TTATCAGACT ACGATGCTAG	5640
TCAGTTGACT GCTAATAAAG TCACTTCTGA CTTCTTTGAA AAAGCTGTTG CCCTAGGTGG	5700
TGATGCCAAA CAAGTCTCTA ACTGGCTCCA AGGGGAAGTC GCTCAGTTCT TGAATGCTGA	5760
AGGTAAACA CTGGAACAAA TCGAATTGAC ACCAGAAAAC TTGGTTGAAA TGATTGCCAT	5820

198  
CATCGAAGAC GGTACTATTT CATCTAAGAT TGCCAAGAAA GTCTTTGTCC ATCTAGCTAA 5880  
AAATGGCGGT GGCGCGCGTG AATACGTGGA AAAAGCAGGT ATGGTTCAAA TTTCAGATCC 5940  
AGCTATCTTG ATCCCAATCA TCCACCAAGT CTTTGCCGAT AACGAAGCTG CTGTTGCCGA 6000  
CTTCAAGTCA GGCAAACGTA ACGCCGACAA GGCTTTACAG GATTCTTAT GAAGGCAACC 6060  
AAAGGCCAAG CCAACCCACA AGTTGCCCTT AAATACTTG CACAGGAATT GGCGAAGTTG 6120  
AAAGAAAAC AGACAGAACA AAACCAGCCC TAAGGTTGGT TTTTCTTCT CTACCAACTC 6180  
CCAATAACTA TTTTGGCTTT ATTTCCAGAG TATTTTATGG TAAAATGAAG AGTAATAATA 6240  
TTTATTAAAG AGGTAAAAAC ATGATTGAAG CAAGTACCTT AAAAGCTGGT ATGACCTTTG 6300  
AAACAGCTGA CGGCAAATTG ATTCGCGTTT TGGAAGCTAG TCACCACAAA CCAGGTAAAG 6360  
GAAACACGAT CATGCGTATG AAATTGCGTG ATGTCCGTAC TGGTTCTACA TTTGACACAA 6420  
GCTACCGTCC AGAGGAAAAA TTTGAACAAG CTATTATCGA GACTGTCCCA GCTCAATACT 6480  
TGTACAAAAT GGATGACACA GCATACTTCA TGAATACAGA AACTTATGAC CAATACGAAA 6540  
TCCCTGTAGT CAATGTTGAA AACGAATTGC TTTACATCCT TGAAAACCTCT GATGTGAAAA 6600  
TCCAATTCTA CGGAACGTAA GTGATCGGTG TCACCGTTCC TACTACTGTT GAGTTGACAG 6660  
TTGCTGAAAC TCAACCATCT ATCAAAGGTG CTAAGTTTAC AGGTTCTGGT AAACCAGCAA 6720  
CGATGGAAAC TGGACTTGTC GTAAACGTTT CAGACTTCAT CGAAGCAGGA CAAAAACTCG 6780  
TTATCAACAC TGCAGAAGGA ACTTACGTTT CTCGTGCCTA ATCTCTAGAA AGAGGTCATT 6840  
CTATGGGAAT TGAAGAACAA CTGGCGGAAA TCGTTATCGC CCCACGTGTA CTTGAAAAAA 6900  
TCATTGCTAT CGCTACTGCA AAGGTAGAGG GTGTTCACTC TTTTCAAAC AGATCAGTGT 6960  
CTGATACCCT TTCAAAACTT TCACTCGGCC GTGGCATTTA TCTTAAAAAC GTGGACGAAG 7020  
AACTCACAGC AGATATCTAT CTCTACCTTG AGTACGGAGT AAAAGTTCCT AAGGTAGCGG 7080  
TTGCTATCCA GAAAGCTGTC AAAGATGCCG TCCGTAATAT GGCTGATGTA GAACTCGCTG 7140  
CTATCAATAT TCACGTTGCA GGTATCGTCC CAGATAAAAC ACCAAAACCA GAATTGAAAG 7200  
ATCTATTTGA CGAGGACTTC CTCAATGACT AGTCCACTAT TAGAATCTAG ACGCCAACTC 7260  
CGTAAATGCG CTTTTCAAGC TCTCATGAGC CTTGAGTTTG GTACGGATGT CGAAACTGCT 7320  
TGTCGTTTCG CCTATACTCA TGATCGTGAA GATACGGATG TACAACTTCC AGCCTTTTTG 7380  
ATAGACCTCG TTTCTGGTGT TCAAGCTAAA AAGGAAGAAC TAGATAAGCA AATCACTCAG 7440  
CATTTAAAAG CAGGTTGGAC CATTGAACGC TTAACGCTCG TGGAGAGAAA CCTCCTTCGC 7500  
TTGGGAGTCT TTGAAATCAC TTCATTGAC ACTCCTCAGC TGGTTGCTGT TAATGAAGCT 7560  
ATCGAGCTTG CAAAGGACTT CTCGATCAA AAATCTGCCC GTTTTATCAA TGGACTGCTC 7620

AGCCAGTTTG TAACAGAAGA ACAATAAGGC TCTTTGTCAA CTGTAGTGGG TTGAAAAAAA	7680
GCTAAGCTCG AGAAAGGACA AATTTCTGCC TTTCTTTTTT GATGTTCAAA GCGATAAAAA	7740
TCCGTTTTTT GAAGTTTTCA AAGTTTCGAA AACCAAAGGC ATTGCGCTTG ATAAGTTTGA	7800
TGAGATTATT GGTGCTTCC AGTTTGGCAT TAGAATAGTG TAGTTGAAGG GCGTTGACAA	7860
TCTTTTCTTT ATCTTTGAGG AAGGTTTTAA AGACAGTCTG AAAAATAGGA TGAGCCTGCT	7920
TAAGATTGTC CTCAATAAGT CCGAAAAATT TCTCTGGTTC CTTATTCTGG AAGTGAAACA	7980
GCAAGAGCTG ATAGAGCTGA TAGTGGTGTT TCAAGTCTTG TGAATGGCTC AAAAGCTTGT	8040
CTAAATCTC TTTATTGGTT AAGTGCATAC GAAAAGTAGG ACCATAAAAT CGCTTATCAC	8100
TCAGTCTACG GCTATCCTGT TGAATGAGTT TCCAGTAGCG CTTGATATCC TTGTATTAT	8160
GGGATTTTCG ATGAAACTGA TTCATGATTT GGACACGCAC ACGACTCATG GCACGGCTAA	8220
GATGTTGTAC AATGTGAAAG CGATCAAGAA CGATTTTAGC ATTCCGGAGT GAAACAGTCT	8280
GGGAGACTGT TTCAGCCTGA GCCTAGGAAT TTGAAAGCGA AGCTGTTTAG CCAAGTCATA	8340
GTAAGGGCTA AACATA?CCA TAGTAATAAT TTTGACGCGA CATCGGACAA CTCTATCGTA	8400
GCGAAGAAAG TGATTTCCAA TGATAGCTTG TGTCTACCC TCAAGAACAG TGATGATATT	8460
GAGATTGTTA AAATCTTGCG CAATGAAGCT CATCTTTCCC TTTGTAAGG CATACTCATC	8520
CCAAGACATA ATCTCAGGAA GACAAGAAAA ATCATGTTTA AAGTGAAAT CATTGAGCTT	8580
ACGAATAACA GTTGAAGTTG AGATGGAAAG CTGATGGGCA ATATCAGTCA TAGAAATCTT	8640
TTCAATCAAC TTTTGAGCAA TCTTTTGGTT GATGATACGA GGGATTTGGT GATTTTCTT	8700
GACGATAGAA GTTTCAGCGA CCATCATTTT TGAACAGTGA TAGCACTGA ATCGACGCTT	8760
TCTAAGGAGA ATTCTAGTAG GCATACCACT CGTTTCAAGA TAAGGAATTT TAGAAGCTTT	8820
TTGAAAGTCA TATTTCTTCA ATTGGTTTCC GCACTCAGGG CAAGATGGGG CGTCGTAGTC	8880
CAGTTTGGCG ATGATTTTCT TGTGTGTATC CTTATTGATG ATGTCTAAAA TCTGGATATT	8940
AGGGTCTTTA ATGTCTAGTA ATTTTGTGAT AAAATGTAAT TGTCCATAT GAATCTTTCT	9000
AATGAGTTGT TTGTCGCTT TTCATTATAG GTCATATGGG ACTTTTTTTC TACAATAAAA	9060
TAGGCTCCAT AATATCTATA GGGGATTAC CCACTACAAA TATTATAGAG CCAACAATAA	9120
AAAGAAAAAG TGTGATAG ATATCAAACA CTTTTTCTT TGCCTCCAC TATCTAAAAA	9180
AATGATAATA GATATAATTG TAAACAAAAA TCCAGATAGG TTTGTCATGA TTGAGAAAGT	9240
TAAAAAACT ATGGCAGAGA ATCGTTAATC TCAGATTGTC GGTAGAACGA TAAACAAGGG	9300
CAAAAAAGAA ACCAATCAGA CTATAATATA ATAACTAAT TGGATCTCTG TGAGATAGTA	9360

200

TCAAATGGCT AATCCCAAAG ATGATAGCAG ATAGGATAAC ATCCAAATAG TACTTGGA	9420
AGGGAAAGAA GGTATTCATA AAATACCCCTC TATCAAGAGT CTCCTCAAAA ACAGGACCGA	9480
TGATTACAGG CAGGACAAAA GATAAGATAG TCGATAAAAA GGTGGTGTGT CCATTTGAAA	9540
AAAGCACGGT AAAATACTCA TCATGAATAT TCCTATGATT AATCAAATGA GCATAGCGTG	9600
CCCAAAAATT ACCGAGAAATC TGATAAACCA CATAAGTTGC AAATAAGTAG AAGACAAATG	9660
ACCAGTTCCA GCTCTTTTTT TCAAAGATAA AGAGCATCTT TTTCTTTTTT AACCTCCAAA	9720
TTAATAGAAG GAAACTTCCC ACTAATCCCA TTGTTAAAAA AAGAGAATAG ACATCAGCTC	9780
CTAACCTTAA AATGATCGTC ACATACAATC CAATTGTTTG TGGTAAATAG GTAGATAGTA	9840
AAATAATAAG CAAAAATAT? CCAAATTGTC TTAGTTTTTT TGTGTTTCTC ATCGTACTTT	9900
TTTGAAAGAT TACCCTGCTC GGAAGCCGTA CTTCCAAGCA TCTATATAAG AATTAAGTGC	9960
CCCTTGCTTC ATATAGGGAG CAAATTCTCT ATAATATAAC CATCTACTAT ATCCATCTTC	10020
CCAAACAGCA AGACCACCTG AAGTTTGCTC CAAGTCCTCA GTTGAAAGAA CTGTAAATGT	10080
ATTTGTACCT GTCATTGCAA GTACCTTCTT AAAATAGATT GTTGTAGGCT CACATTTATA	10140
GTATATTTCT TTTTGTCT ATTTTATAGC CCATCTCCTC AACTGGCAAT TTTTCGACCT	10200
GAATTACATT TTTCCATAAA AAATGAGACC TTTCTAGTCT CATTAGTCA TTCTTAGTAT	10260
TTTCTAAATC GTTGATAGCG TTCTTCCAGC AACTCTTCTA GCGGTTTTTG TGAAAGTCTA	10320
GCCAGCTCCG TTTGGAGTTC TTTTTTGACA CTCTTAATCA GTTCTTTACT AGAAAGTCTT	10380
ATTCAGAAA TCACCTTATC CACCACGTCC ATTTCTAACA GTTCATGCGA AGTGATTTTC	10440
ATCAGTTCTG CTGCTTCCAT AGCGCGAGTA CCGTCCTTCC ATAAAATGGA AGCAAAGCCT	10500
TCTGGACTGA GAATGGCATA GATAGAATTT TCCAGCATCC AGACACGGTC CGCGACAGCT	10560
AGAGCCAGAG CCCCCCTGA ACCACCTTCA CCGATAATAA TGGCGATAAT AGGAACTTTC	10620
AGGTCACTCA TTTCCATGAG ATTGCGAGCG ATAGCTTCCC CTTGACCACG TTCTTCCGCT	10680
CCGACACCAG GATAAGCACC TGCTGTATTG ATAAAGGTCA CAACTGGACG GCCAAATTC	10740
TCAGCCTGTT TCATCAACCG CAGTGCCTTT CGGTAGCCTT CTGGATGTGG TTGGCCAAAA	10800
TTCCGTTTGA GGTGTCTTG CAAACTCTTG CCTTTTTGGA TACCAACCAC TGTACAGCT	10860
TGCTCTCCAA GCCAACCAAT ACCACCAACA ACTGCACCAT CATCACGAAA AGAACGGTCA	10920
CCATGTAATT GGATAAATTC ATCAAAAATG CCTGTCGCAA AGTCCAAGGT TGTCAAGCGA	10980
CTCTGCTCAC GCGCTTCTCT GACTATTTTT GCAATATTCA TCTAGGACTC CCTCCATGCA	11040
ATCTGACTAG GCTAGCAATC GTATCTGGTA AGTCTCTTCT TTTGACAATA GCATCCACAA	11100
AGCCATGTTT TAATAGGAAT TCTGCCTTTT GGAATCCTC AGGCAAGCTT TCACGAACCG	11160

201

TATTTTCAAT CACACGACGC CCAGCAAAAC CAACEAAGCT CTGTGGTTCA GCCAGAATGA	11220
TATCGCCTTC CATAGCGAAA GAAGCTGTCA CACCACCAGT CGTTGGATCT GTCAAAATGG	11280
TCAGGTAAAA GAGACCAGCA TTTGAATGGC GTTTAACCGC CGCAGAGATC TTAGCCATCT	11340
GCATGAGACT CATGATTCCT TCCTGCATAC GGGCTCCACC AGAGGCTGTG AATAGGACAA	11400
CTGGCAATTT TTCGACAGTC GCATACTCAA ACAAACGAGT GATTTTTTCA CCTACAACCG	11460
TACCCATAGA AGCCATGATA AAGTTAGAAT CCATAATCCC AAGAGCCACA GTCTGACCTT	11520
TAATAAGAGC AGTTCCTGTC ACAACGGCTT CATGCAGACC TGTTTTTTCA CGCATAGATG	11580
CCAGTTTCTT TTGGTAACCA GGGAAATGCA AGGGATCCTT GCTTTCAATC CCTGTAAACA	11640
ATTCTTTGAA GGTTCCTATA TCAATCGTCA AAGCCAAGCG TTCTTGGGCA GAAATACGAA	11700
AGGTATAGCT ACAGTGGCGA CAGATACGTT CACTTCCCAG ATCCTTCTGA TAGATGGTAT	11760
GCTTACAGCC TGGACACTGG GAAAATAATT CATCTGGAAC CTCTGGCTTA GCTTGAGGTT	11820
TTCCCTAAC CGAACGATTG GGATTGATTC GAATATACTT ATCTTTTTTA CTAAATAGAG	11880
CCATTGATTC CCCTTTTCGG TTTAACTCT TAAAGTCATT TTATTCTTTT TCTTGATATT	11940
TAGGTAAGAA GGTTCCTATC AAGAAGGAAG TATCATAATC CCCAGCAATG ACATTGCGAT	12000
CTGAAATGAG GTCAAGCTGG AAATCTGCAT TGGTCTGCAC TCCTTCAATT TCTAATTCAT	12060
AGAGGGCAGC TTGCATTTTC ATCAAGGCGT CAAAACGATT TTGGCCCTGT ACTATGATTT	12120
TGGCAATCAT ACTATCATAA TAAGGCGGAA TGGTATAACC TGGATAAACT GCTGAATCCA	12180
CGCGCAAGCC AACTCCACCA CTTGGCAGAT AGAGATTAGT AATCTTACCT GGACTTGGAG	12240
CAAAGTTAAA GGCTGGGTTT TCTGCATTGA TACGACACTC GATGGCATGA CCGCGTAGGA	12300
CAATATCTTC TTGCTTAACA GACAAAGGCT GACCTGCCGC AATGCAAATC TGTTCCCTAA	12360
CGATATCAAC ACCTGAAACA AACTCTGTTA CTGGATCTTC TACCTGAACA CGAGTATTCA	12420
TCTCCATGAA ATAGAAATTG CTAATTGCTT CATCAAGAAG AAATTCAATG GTTCCTGCAT	12480
TCTCATAGCC AACAACTCT GCCGCTCGAA CAGCAGCAGC ACCTATTTCA TGACGCAGCG	12540
TTTTTCCGAT TGCAATCGAG GGACTTTCTT CCAAAACCTT TTGGTTATTC CTTTGAAGAG	12600
AACAATCCCG TTCACCCAAG TGAATCACAT GTCCATGCTC ATCACCTAGG ATTTGAACCT	12660
CAATGTGCCG AGCTGGATAG ATAACCCGTT CTATGTACAT GGCACCATTG CCATAATTGG	12720
CCTTGGCCTC ACTAGAGGCA GTTTCAAAGG CAGAAACGAG GTCATCTGGT TTTTCAACCT	12780
TACGAATCCC TTTACCACCT CCACCTGCTG AAGCCTTGAG CATAACAGGA TAGCCAATTT	12840
TTTCAGCAAC AATCAAAGCT TCTTCAGAGT TATGCACTTC TCCATCTGAA CCTGGTATAA	12900



202

CAGGCACACC	TGCTTTAATC	ATCTGAGCAC	GCGCATTGAT	CTTATCCCCC	ATCATATCCA	12960
TAACATGACC	AGATGGACCG	ATAAACTTGA	TACCTACTTC	TTCACACATG	GTCGCAAATT	13020
TGGAATTTTC	ACTGAGAAAT	CCAAAACCAG	GGTGAATAGC	TTCTGCCTCA	GTCAAGACTG	13080
CAGCTGATAG	AACTGCATTA	ATATTGAGAT	AAGACTCTGT	TGCCTTGCCA	GGACCAATAC	13140
AAACTGCTTC	ATCTGCCAAA	AGCGTATGAA	GAGCTTCCTT	ATCAGCAGTT	GAATAAACCG	13200
CTACCGTCGC	AATCCCAAT	TCACGTGCCG	CACGGATAAT	ACGAACCGCA	ATTCACCAC	13260
GATTGGCAAT	TAAAATTTTT	CGAAACATGG	AGAACCTCCT	TAGTTCCCAA	TTGCAAAAGT	13320
AAGGTACCA	CTGGCTGCAA	GCTTGCCATC	CACCTCAGCC	TTTGCTTCAA	CCACAGCTAT	13380
GGTGCCACGA	CGTTTTACAA	AAGTCGCTGT	CATAACCAAT	TGGTCGCCTG	GTACAACTTG	13440
CTTCTGTAAC	TTAACCTTGT	CCATACCAGC	GTAAAAGACC	AGTTTTCTTT	TATTTTCAGG	13500
TTTTGATAAC	TCCAACACAC	CGGCAGTTTG	CGCCAAGGCT	TCCATAATCA	CAACACCTGG	13560
CATAACTGGG	TATTGAGGAA	AGTGGCCGTT	AAAGAAAGGC	TCGTTGATGG	TCACATTTTT	13620
GATAGCAACA	ATGGTATCCT	CGCTCACTTC	CAAGACACGG	TCCACTAGAA	GCATAGGATA	13680
ACGGTGGGGA	AGAGCTTCTT	TGATTCTTGG	AATATCGATC	ATTTGATACG	TACCAATCCT	13740
TTACCAAAC	CAACCATTTC	TTCGTTAGAG	ACGAGAATTT	CCGTTACCAC	ACCATCCTTA	13800
GGAGCTGGGA	TTTCATTTCAT	GACTTTTCATG	GCTTCGATAA	TTACCAATGT	TTGACCTTTT	13860
TTGACACTAT	CACCAACTGT	AACGAAGGCA	GGTTTTATCTG	GTCCAGCAGC	CAAGTAAACC	13920
ACTCCAACAA	GTGGACTCTC	TACAAGATTT	CCCTCAGTAG	CCACACTTGC	TTCAGCTGGA	13980
GCTGGAAC	CTTCTGCTAC	AGTCTCTGCT	GGAGCAGATG	TAGGAGCTAC	TGGACTCGGT	14040
GTTGCTAGAA	CGGGTGCTGG	AGCGACTTGA	GTTGCAACTT	CAGGCACAGG	TCTTGCTTCA	14100
TTCTTGCTAA	ACTGCAACTC	ATCCGTCCCA	TTTTTATAAG	AAAATCTCT	CAAACTTGAC	14160
TGGTCAAATT	GAGTCATCAA	GTCTTTAATA	TGTTTTAAAT	TCATACTTAT	CTATTCTCCC	14220
AACGTTTGAA	AGCAAGAACT	GCATTGTGGC	CTCCAAAACC	AAAAGTATTT	GAAATAGCGT	14280
ATGGAATTT	TTTCTCCAAG	CCTTGTCCTAT	AAACGACATT	AGCTTCGATA	TAATCTGATA	14340
CTTCACTTGT	CCCAGCTGTC	ATTGGTACAA	AGTTATGACG	CATAGCTTCG	ATGGTGACGA	14400
TAGCTTCTAC	TGCACCCGCA	GCCCCCAGCA	AATGTCTCTGT	AAAAGACTTG	GTTGATGATA	14460
CAGGTACTTC	CTTACCAAGA	ACAGCTACGA	TAGCACCAC	TTCTCTTTT	TCATTGGCAG	14520
GAGTTGACGT	TCCGTGAGCA	TTGACATAGG	CTACTTGCTC	TGGAGAAATC	TCAGCTTCTT	14580
CCAAGGCTAG	TTTGATGGCC	TTGATAGCTC	CCTGACCTTC	TGGATGTGGA	GAAGTCATGT	14640
GGTAGGCATC	ACAAGTATTT	CCGTAACCAA	CCACTTCAGC	CAGGATAGTA	GCTCCACGTT	14700

203

TTTCAGCGTG	TTCAAGACTT	TCTAGAACCA	ACATCCCTGA	ACCTTCACCC	ATAACAAACC	14760
CATTGCGATC	CTTATCAAAT	GGGATCGAAG	CACGAGTTGG	ATCCTCTGTA	GTAGAGAGAG	14820
CTGTTAAGGC	TTGGAACCA	CGGATGGCAA	AAGGTGTGAT	AGAAGCTTCT	GTTCTCTCCA	14880
CCAACATCAC	ATCTTGAAA	CCAACTTAA	TGGAGCGGAA	GGCATCCCCA	ATCGCATCAT	14940
TTGATGAAGA	GCAGGCAGTA	TTGATAGATT	TACAAACACC	GTTTGCACCA	AAACGCATGG	15000
CTACATTCCC	AGAAGCCATA	TTTGGTAAAG	CTTTTGAAG	AGTCATTGGT	TTGACACGTT	15060
TGGGTCCTTT	TTCATGAAGG	CGAAGTACCT	GATCTTCAAT	TTCTTGATT	CCACCAATAC	15120
CAGATGCAAC	GATAACACCA	AAACGATCCC	TATTAAGAGC	CTCTACATCA	AGATTGGCAT	15180
GATTTACAGC	CTCTGGGCT	GCATACAAGG	CATATAAAGA	ATAGTTATCA	AAACGGTTGG	15240
TATCTTTTTT	TACAAAGTAT	TTATCGAAGC	GAAAACTTG	GATTCTGCC	GCATTATGCA	15300
CATCAAAGTC	ACTATGATCA	AATTTTGTA	TGCCACCAAT	GCCGATTTTC	CCAGTTGCTA	15360
AACTATTCCA	AAATTCTTCT	GGTGTATTTT	CGATTGGAGA	TGTTACTCCA	TAACCTGTTA	15420
CCACTACTCG	ATTTAGTTTC	ATTCTTTTCA	CCTCTAGCTT	TCGCTACATA	CTTAAGCCAC	15480
CATCAATGGC	AACCACTTGT	CCAGTTAGAT	AATCTTGGCC	TGCTAAAAAT	ACTGTCAAAT	15540
CTGCAACCTG	CTCTGCCTGC	CCAAATTCTT	TCATCGGAAT	CTGAGCTAGT	GTAGCTTCCT	15600
TAATCTTATC	TGACAGGATA	GCGGTCATAT	CAGACTCAAT	CATTCTCGGA	GCAATCACAT	15660
TGACTCGTAT	ATTCCGACTA	GCGACCTCGC	GTGCCACAGA	CTTGGTAAAG	CCAATCAAGC	15720
CAGCCTTAGA	AGCAGCATAA	TTAGCTTGAC	CAATATTTCC	CATCAAACCA	ACAACACTAG	15780
ACATATTAAT	GATAGCACCT	TCTCTGGCTT	TCATCATCGG	TTTCAAGACT	GATTGTGTCA	15840
TATTAAGGC	ACCAGTCAGA	TTGACCTTGA	GCACTTTTTT	AAAATCTGCT	TCTGTCTCTT	15900
TGAGCATAAG	AGTATCTTGG	GTAATCCCTG	CATTGTTGAC	CAAAACATCT	ACTGAACCCA	15960
GTTCTGCAAT	AGCTTGATCA	ATCATACGCT	TAGCGTCTGC	AAAATCTGAT	ACATCTCCTG	16020
AAATGGGAAC	CACCTTGATA	CCATAGTTTG	AAAATCAGC	GAGCAATTCT	TCTGAGATTG	16080
CCCCACGACT	GTTTAAGACA	ATGTTGGCTC	CTGCTTGAGC	AACTTGTGG	GCGATGGCAA	16140
GACCAATTCC	ACGACTCGAA	CCTGTAATAA	AGATATTTTT	ATGTTCTAGT	TTTATTTTTT	16200
TCCTTTCAAA	ACTTCTACTT	ATTTTAGTCT	ATTTTCTTAA	AAGTGCTACT	AACTCGCTT	16260
GATCTTCCAC	ATGAGCTAAG	TGAGCAGTTT	GATCAATTTT	TTTAACAAAA	CCTGACAAGA	16320
CTTCCCCCGG	TCCAATCTCG	ATAAAGTTGC	TTATGCCTGC	TTCTTGCATG	ACCCCAATAC	16380
TTTCATAGAA	ACGAACGGGT	TCCTTGACCT	GACGCGTCAA	GAGCTGAGCA	ATGTCTCTTT	16440

204

TTTGCAACAC AGCAGCTTCT GTATTGCCGA CTAGGGGACA AGTAAATCT GAAAACTTA	16500
CCTGAGCTAG AGTTTCAGCT AGTTTCTGGC TAGCAGGTTT AAGGAGAGCG GTGTGAAAGG	16560
GACCTGACAC CTTAAGAGGA ATCAAGCGTT TGGCACCTGC TTCTTGCAAA AGTTCAACCG	16620
CTCGATCAAC TGCAACCACT TCTCCAGCAA TGACGATTTG TGCAGGTGTG TTATAGTTGG	16680
CTGGAGTAAC CACTCCAAGT TCAGAAGCTT TTTGACAGGC TTCTTCAATG ACCTCTACTG	16740
GCGTATTGAG AACTGCTACC ATCTTGCCAG AGTCAGCAGG AGCCGCTTCT TCCATATAGG	16800
CTCCACGCTT AGCTACCAAG GCAACCGCAT CTTCAAAATC CAAGGCGCCA CTTGCCACCA	16860
AGGCAGAGTA TTCTCCAAGA GACAAACCAG CAACCATATC AGGCTGATAG CCTTTTCTT	16920
GCAATAAACG GTAGATAGCA ACCGAAGTCG CTAGAATGGC TGGTTGCGTA TAGCGGGTCT	16980
GATTGAGTTT GTCTTCTTCC GTATCGATGA GATAACGCAA ATCATAACCG AGCACCTGGC	17040
TCGCTCGATC AATCGTTTCT TTAACAATCG GATACTGATC ATAGAAATCC CGTCCCATCC	17100
CTAGATACTG GGCACCTTGA CCAGCAAATA AAAAGGCTGT TTTAGTCATT TCTTACAACT	17160
CCTGTCCAGC GAGAGGCTTC TTCTTGAATT TTCTTAGCGG CTCCGTAATA CAAATCTTTT	17220
AGGATTTCTT CAGCTGTTTC TTCTTTAGAA ACAAGCCCTG CGATTTGACC TGCCATAACA	17280
GAGCCACCAT CCACATCACC GTGAACAACT GCTTTGGCTA GAGCACCTGC TCCCATTGT	17340
TCAAAGATTT CTAATCAGG ATCTTCTTGC TTAAAGGCAT CTTTTTCAGC CAGTTCAAAA	17400
TCTCTAGTCA ACTGATTTTT AATAGCACGA ACAGCATGAC CAAAGTGCTG AGCTGAAATC	17460
GTAGTATCAA TATCCCTTGC TTTTAAAATT TTCTCCTTGT AGTTTGGATG GGCATTCGAC	17520
TCTTTTGCAA CTACAAACCG TGTCCCCACC TGTACAGCCT CTGCACCTAG CATAAAGCCA	17580
GCCGCAGCAC CTTCAACCAT CGCAATTCCT CCTGCAGCAA TAACAGGAAT AGATATAGCT	17640
GTGGCTACCT GTCGCACCAA GGTCAATGGT GTTAATTTAC CGATATGCCC CCCAGCTTCC	17700
ATTCCTTCTG CAATAACAGC GTCTGCACCG ATTTTTTCCA TCGGTTTAGC TAAAGCGACA	17760
CTAGGAACAA CAGGAATAAC GATTATCCCA GCTTCATGGA AACGTTCCAT ATACTTGCTT	17820
GGATTTCTG CTCCTGTTGT GACAACCTTA ACACCTTCTT CAATAACGAG ATCCACGATG	17880
TCTTCCACAA AGGGAGATAA GAGCATGATG TTGACCCCAA AGGGTTTATC AGTCAATGAT	17940
TTGATTTTAT CAATATTGGC CTTGACAACT TCTTTCGGG CATTTCCTCC ACCGATAATT	18000
CCTAATCCTC CAGCCTTGA AACAGCCCTT GCCAAATCAC CATCAGCAAC CCAGGCCATC	18060
CCTCCTTGA AAATAGGATA ATCAATCTTC AATAATTCTG TAATACGCGT TTTCATAGTG	18120
CCTCCAACCT TCCTTGCTTA CGTAATAGTT CGATTTTACC ATAATTTGAC AGTCAAACTA	18180
TTACCTAAAC AAGAGGGAGT GGGTTTCTCC CTACTCCTTC TACTAATATT CTGCTTATTT	18240

205

TGCTTGCTCT TCAACGTAAG CAACCAAGTC ACCAACTGTT TTCAAGTCAT TTTCTGCTTC	18300
GATTTGGATA TCAAAAGCAT CTTCCGATTTC TGAGATTACT TGGAAACAAGT CCAATGAATC	18360
TGCGTCCAAA TCATCAAAAG TTGATTCAAG TGTACTTCTT GATGCGTCTT TTCCAAGTTC	18420
TTCAACGATA ATTTCTTGTA CTTTTTCAAA TACTGCCATG ATAGGACTCC TTTAAATAA	18480
ATAGTTTTTT TATAACAATG TGTTCAACCAC ATGATTACCT AAATTGTAAG AATGAGCGTG	18540
CCCCAGGTCA AGCCTCCACC GAAGCCTGAT AGAAGAACAG TCTGGCTACC ATCTAAAGGG	18600
ATGAGACCTT GTTCTACACA CTCTGAAAGT AAAATCGGGA TACTGGCTGC ACTGGTATTG	18660
CCATATTCCA TCATATTGGC TGGAAAGTTG GCTCGGTCAA CACCAATTTT TCTAGCCATC	18720
TTATCCAAAA TACGGTCATT GGCTTGATGA AGTAGCAGAT AATCCAAGTC TGTCACCTCT	18780
ATAGGAGATT CATCAATAGT CTGCTTGATA GACTTGGCTA CATCTCGAAT GGCAAAATCA	18840
AAGACTGTGC GTCCATCCAT CTTCAAAAAC GAATCTGCAC TTTCTTGATC TGAAAATGGA	18900
GAATGTAAAC CTGAATGCCC ATAAGTTAAA CACTCGCTGC GACTTCCATC GCTATTGAGA	18960
CTCTCAGCTA AGAAATGCTC TTGCTCGCTA GCTTCTAACA AGACACCACC AGCACCATCT	19020
CCAAACAACA CAGCTGTGTA TCGATCCGAC CAATCGACTG CCTTAGAGAG GGTTTCACTA	19080
CCAATCACC AAGCTTTTGG AAAGCGACCA GAAGCGATAA ACTTTTCAGC AGTTGAAAGA	19140
GCAAAATACAA ATCCACTGCA AGCCGCGGTT AAGTCAAAAG CAAAGGCTTT ATTAGCACCA	19200
ATATTAGCTT GAACACGAGC AGCTGTAGAG GGCATCATCG AATCTGGAGT AATGGTAGCT	19260
AGGATGATAA AATCCAGTTC TTCTCCTGTT ATTCCAGCTT TTGCCATCAG TTTCTTAGCA	19320
ACCTCTGTAG CCAAATCACT CGTAGATTCT GTTCTTGAAA TATGCCCTTG TCGTATTCCC	19380
GTTCCGACTG AAATCCACTC ATCATTGGTA TCCATAATCT GAGCCAAGTC GTGATTTGTA	19440
ACCACTTGCT CTGGCACATA ATGAGCAACC TGACTTATTT TTGCAAAAGC CATTATTTCA	19500
AATCCTCCAA AAATTGGTAA AGATTAGTCA AACCTTTACC CATGACAGCA ATTTCTTCCT	19560
CGCTCATGCC ATCAATAATT TTTTCTACCA TGGCCTTGTC GAAGCGTTTA TGCAGTCTAT	19620
GAATCAAGCG ACCCTTCTTT GTCAAATGCA GATGCACCAC ACGACGATCC TGTCTTGACC	19680
GAACTCGCTC AATGTAGCCC GG	19702

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5211 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GAAAAATTCC TCTCTTCTCT TGAAAAATTT TGAaaaaaATG GTATGATAGT AACAAAGTTAT	60
TTTTAAGAGG AAAGAAAGGG GAATAATGGA GAAAATCAGT TTAGAATCTC CTAAGACGGG	120
GTCGGACCTA GTTTTGGAaA CACTTCGTGA TTTAGGAGTT GATACCATCT TTGGTTATCC	180
TGGTGGTGCG GTTTTGCCCTT TTTATGATGC GATATATAAT TTAAAGGCA TTCGCCACAT	240
TCTAGGGCGC CATGAGCAAG GTTGTTTGCA TGAAGCTGAA GGTATGCCA AATCAACTGG	300
AAAGTTGGGT GTTGCCGTCG TCACTAGTGG ACCAGGAGCA ACAAATGCCA TTACAGGGAT	360
TGCGGATGCC ATGAGCGATA GCGTCCCCCT TTTGGTCTTT ACAGGTCAGG TGGCCGAGC	420
AGGGATTGGG AAGGATGCCT TTCAGGAGGC AGACATCGTG GGAATTACCA TGCCAATCAC	480
TAAGTACAAT TACCAAGTTC GTGAGACAGC TGATATTCCG CGTATCATTG CGGAAGCTGT	540
CCATATCGCA ACTACAGGCC GTCCAGGGCC AGTTGTAATT GACCTACCAA AAGACATATC	600
TGCTTTAGAA ACAGACTTCA TTTATTCACC AGAAGTGAAT TTACCAAGTT ATCAGCCGAC	660
TCTTGAGCCG AATGATATGC AAATCAAGAA AATCTGAAG CAATTGTCCA AGGCTAAAAA	720
GCCAGTCTTG TTAGCTGGTG GTGGAATTAG TTATGCTGAG GCTGCTACGG AACTAAATGA	780
ATTTGCAGAA CGCTATCAAA TTCCAGTGGT AACCAGTCTT TTGGGACAAG GAACGATTGC	840
AACGAGTCAC CCACTCTTTC TTGGAATGGG AGGCATGCAC GGGTCATTCT CAGCAAATAT	900
TGCTATGACG GAAGCGGACT TTATGATTAG TATTGGTTCT CGTTTCGATG ACCGTTTGAC	960
GGGGAATCCT AAGACTTTCG CTAAGAATGC TAAGGTTGCC CACATTGATA TTGACCCAGC	1020
TGAGATTGGC AAGATTATCA GTGCAGACAT TCCTGTAGTT GGAGATGCTA AGAAGGCCTT	1080
GCAAATGTTG CTAGCAGAAC CAACAGTTCA CAACAACACT GAAAAGTGGG TTGAGAAAGT	1140
CACTAAAGAC AAGAATCGTG TTCGTTCTTA TGATAAGAAA GAGCGTGTGG TTCAACCGCA	1200
AGCAGTTATT GAACGAATTG GTGAATTGAC GAATGGAGAT GCCATTGTGG TAACAGACCT	1260
TGGTCAACAC CAAATGTGGA CAGCTCAGTA TTATCCCTAC CAAAATGAAC GTCAGTTAGT	1320
GACTTCAGGT GGTTTGGGAA CAATGGGCTT TGGAAATCCA GCAGCAATCG GTGCTAAAAA	1380
TGCTAACCCA GATAAGGAAG TAGTCTTGTT TGTGGGGAT GGTGGTTTCC AAATGACCAA	1440
CCAGGAGTTG GCTATTTTGA ATATTTACAA GGTGCCAATC AAGGTGGTTA TGCTGAACAA	1500
TCATTCACTT GGAATGGTTC GCCAGTGGCA GGAATCCTTC TATGAAGGCA GAACATCAGA	1560
GTCGGTCTTT GATACCCCTC CTGATTTCCA ATTGATGGCG CAGGCTTATG GTATTAAAAA	1620
CTATAAGTTT GACAATCCTG AGACCTTGGC TCAAGACCTT GAAGTCATCA CTGAGGATGT	1680

TCCTATGCTA ATTGAGGTAG ATATTTCTCG TAAGGAACAG GTGTTACCAA TGGTACCGGC	1740
TGGTAAGAGT AATCATGAGA TGTGTTTTGGT GCAGTTCCAT GCGTAGAATG TTAACAGCAA	1800
AACTACAAAA TCGTTCAGGA GTCCTCAATC GCTTTACAGG TGTCTATCT CGTCGTCAGG	1860
TTAATATTGA AAGCATCTCT GTTGGAGCAA CAGAAGATCC GAATGTATCG CGTATCACTA	1920
TTATTATTGA TGTGCTTCT CATGATGAAG TGGAGCAAAT CATCAAACAG CTCAAATCGTC	1980
AGATTGATGT GATTGCGATT CGAGATATTA CAGACAAGCC TCATTTGGAG CGCGAGGTGA	2040
TTTTGGTTAA GATGTCAGCG CCAGCTGAGA AGAGAGCTGA GATTTTAGCG ATTATTCAAC	2100
CTTTCCGTGC AACAGTAGTA GACGTAGCGC CAAGCTCGAT TACCATTGAG ATGACGGGAA	2160
ATGCAGAAAA GAGCGAAGCC CTATTGCGAG TCATTGCGCC ATACGGTATT CGCAATATTG	2220
CTCGAACGGG TGCAACTGGA TTTACCCGCG ATTAAAAATC CAACTTAAAT TTATTAAACC	2280
AGCCTAAAAG GCAATAAATA ATAGAAAAGA GAGAAAAGCT ATGACAGTTC AAATGGAATA	2340
TGAAAAAGAT GTTAAAGTAG CAGCACTTGA CGGTAAAAAA ATCGCCGTTA TCGGTTATGG	2400
TTTACAAGGG CATGCGCATG CTCAAAACTT GCGTGATTCA GGTGCTGACG TTATTATCGG	2460
TGTACGTCCA GGTAAATCTT TTGATAAAGC AAAAGAAGAT GGATTTGATA CTTACACAGT	2520
AGCAGAAGCT ACTAAGTTGG CTGATGTTAT CATGATCTTG GCGCCAGACG AAATTCAACA	2580
AGAATTGTAC GAAGCAGAAA TCGCTCCAAA CTTGGAAGCT GGAAACGCAG TTGGATTTGC	2640
CCATGGTTTC AACATCCACT TTGAATTTAT CAAAGTTTCT CCGGATGTAG ATGCTCTCAT	2700
GTCTGCTCCT AAAGGACCAG GACACTTGGT ACGTCGTACT TACGAAGAAG GATTTGGTGT	2760
TCCAGCTCTT TATGCAGTAT ACCAAGATGC AACAGGAAAT GCTAAAAACA TTGCTATGGA	2820
CTGGTGTAAG GGTGTTGGAG CGGCTCGTGT AGGTCTTCTT GAAACAACTT ACAAAGAAGA	2880
AACTGAAGAA GATTTGTTTG GTGAACAAGC TGTACTTTGT GGTGGTTTGA CTGCCCTTAT	2940
CGAAGCAGGT TTCGAAGTCT TGACAGAAGC AGGTACGCT CCAGAATTGG CTTACTTTGA	3000
AGTTCTTCAC GAAATGAAAT TGATCGTTGA CTTGATCTAC GAAGGTGGAT TCAAGAAAAT	3060
GCGTCAATCT ATTTCAAACA CTGCTGAATA CGGTGACTAT GTATCAGGTC CACGTGTAAT	3120
CACTGAACAA GTTAAAGAAA ATATGAAGGC TGTCTTGGCA GACATCCAAA ATGGTAAATT	3180
TGCAATGAC TTTGTAAATG ACTATAAAGC TGGACGTCCA AAATTGACTG CTTACCGTGA	3240
ACAAGCAGCT AACCTTGAAG TTGAAAAAGT TGGTGCAGAA TTGCGTAAAG CAATGCCATT	3300
CGTTGGTAAA AACGACGATG ATGCATTCAA AATCTATAAC TAATTAGAAA TATATAGCCC	3360
TGGAGATGAT TTTATGAAAA AGATTATGAG AAAAATTGCA TCGTTATTAT TGCTTCTAGT	3420

208

TGTATAATGT AATTACACCG TCGGTAATAG TGCTAGCAGA CCAAAATAAA GCAGATTGGT	3480
CGTATGATGA AAATGCTGTA ATTAACATT ATGATGATGC TAAATTTGAA GATGGTAGGT	3540
TGCATATGAA CTTTGAACAA TTCTTCAAAT TGGCACAAAT AGCTAGAGAA GAAGGCTTTG	3600
AAATTCATT CCGTTTGAG AGAGCTGGTG CGACTAAATC TGCTCGTTAT ATAGCGAAAT	3660
GGATTTTGAG AAATAAAAAA CATTAACAAA TATAGTTGGT AAATCATTAG GACCTAAATC	3720
AGCTGTTAGA TTCGGAGAAG CTTTATCCTA TATTGAAGGT CCTCTTCGCA GAATAAATGA	3780
GACGATAGAT GGCGGTTTAT ATCAAATAGA GCAAATTATT GCATCTGGAT TGAAAGAATC	3840
GGGTTTAAAT GACTGGACTG CGAAAACCTT AGCTTCAGCT ATTCTGGGA TATTAGATGT	3900
ACTTATTTAG GGGTTGAAAT CATATGAATA TTACCAATTT GTTTCTATC AAGACAGGAT	3960
GTGATGAAAC TGATAGGCAA CTGCAAAAC TATTTTTCA GTTGGATTGA CAATTGGGAG	4020
AATTGACAGA TCAACTAAGA AAATTAGATT CTAATTTTGT TCCTCGTAGT CAATTTGTAG	4080
ACACGTGGA TTTGAATGAT GTAGAATATA AAGAAATTTT AAATATTTT ATCTTCCATC	4140
GTAATGATAG TGAAGAAAGT TTGGTAGAAT GGTATATGA TTGGATTTC ACAATCGTT	4200
ATGAACCTCC TAAAGACTTT TCGATTCTGA TGGCTCATAA ATACCATGAA AGTGTTACTG	4260
AAGTTTTCGG AGATGAATAA CTAATAAACA GTCATTAGTG ACTGTTTTT ATAGAAAAAG	4320
AGGTTTTATA TGTTAAGTTC AAAAGATATA ATCAAGGCTC ACAAGGCTT GAACGGTGTG	4380
GTGTGAATA CTCCACTGGA TTACGATCAT TATTATCGG AGAAGTATGG TGCTAAGATT	4440
TATTTGAAAA AAGAAATGC CCAGCGTGT CGCTCCTTTA AAATTCGTGG TGCCTATTAT	4500
GCCATTTCCC AGCTCAGCAA GGAAGAACGT GAACGTGGGG TAGTCTGCCG TTCTGCGGGA	4560
AATCATGCGC AGGGAGTAGC CTATCTTGT AATGAAATGA AAATTCCTGC TACTATCTTT	4620
ATGCCCATTA CTACGCCACA ACAAAGATT GGTCAGGTT CTTTTTTGG TGGGGATTTT	4680
GTAACATTA AACTAGTTGG AGATACCTTT GATGCCCTCAG CCAAAGCAGC TCAAGAATTT	4740
ACAGTCTCTG AAAATCGTAC CTTTATTGAT CTTTTGATG ATGCTCATGT TCAAGCAGGT	4800
CAAGGAACAG TTGCTTATGA GATTTTAGAA GAAGCTCGAA AAGAATCGAT TGATTTGAT	4860
GCTGTCTTGG TTCCTGTTGG TGGTGGCGGT CTCATTGCCG GGGTTTCTAC CTATATCAAG	4920
GAAACAAGTC CAGAGATTGA GGTATCGGA GTAGAGGCGA ATGGAGCGCG TTCCATGAAA	4980
GCTGCCCTTG AGGCTGGAGG TCCAGTAAAA CTCAAGGAAA TTGATAAATT TGCTGATGGG	5040
ATTGCTGTGC AAAAGGTAGG TCAGTTGACC TATGAAGCAA CTCGTCAACA TATTAAAACT	5100
TTGGTAGGTG TCGATGAGGG ATTGATTCT GAAACCTTGA TTGACCTTTA CTCTAAGCAA	5160
GGGATAGTCG CAGAACCTGC TGGAGCGGCT AGTATCGCCT CTTTAGAGGT TTAGCTGAA	5220

209

TATATTAAGG GGAAAACCAT TTGTTGTATC ATTTCTGGAC GAAATAATGA TATCAACCGT 5280  
 ATGCCAGAAA TGGAAGAGCG TGCCTTGATT TATGATGGTA TCAAACATTA CTTTGTGGTC 5340  
 AATTTCCAC AACGTCCAGG AGCTTTGCGT GAGTTTGTA ATGATATCCT GGGGCCAAAT 5400  
 GATGATATCA CACGTTTGA GTATATCAAA CGAGCTAGCA AGGGAACAGG CCCAGTATTA 5460  
 ATTGGGATCG CTTTAGCAGA TAAGCATGAT TATGCAGGTT TGATTCTAG AATGGAAGGT 5520  
 TTTGATCCAG CTTATATTAA CTTAAATGCT AATGAAACGC TTTATAATAT GCTTGTCTGA 5580  
 GGACTAATAA AAAAATATCA TACCTTCATT TTGATTTCTT ATCTATTGAC AAGCATAGTC 5640  
 ACACTGTCTT TAATACTCTT CGAAAATCTC TTCAAACCAC GTTAGCTCTA TCTGCAACCT 5700  
 CAAAACAGTG TTTGAGCAA CTTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTCATTG 5760  
 AGTATAAGGT ATGATTTGAT TCTTTTGTG TGACAAATAT ACTATATTAA AAAGATATAT 5820  
 AAGTAATTAA CTGAGCTTAT CTGTCTTGTC ATCTCTATTA AGGATGGTT AGATAATCGG 5880  
 GTGTCTGCTT CTAGGCTAGC ACCTCAATAT CCAAAGGAGT GATCAATTG AAGGACATAA 5940  
 GGAATACCTA TCTCTCAGAT GATTTATTGA GGAAGAAAGA TAGCAGTTTT TGAGCTAGTG 6000  
 AAGGCTTGGG TTTCTAAAGG TTAGAACTAT CATCTTCAGT TCTTAAATCG AAGAAATAAG 6060  
 CTATCTTACG GAAATAGAGA AGCATTTTTT AAGAACTTGA ATAATTTCCG ACCTTAAGAG 6120  
 GGTAAATAATA CAGTATTTTT ATTAGCAAAT ATTTATGGTG TAGAGGCTAG CAAAACCTAT 6180  
 ATATTATCGG ATTTAAAAAG GAAGTAAGAA A 6211

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7939 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CCGGACTCCC CACGATTCTT CAAAATAACT GAGTATATTT CTATCTTGAT TTTCAGATAT 60  
 AAATCTCTCC TTCTGTGGCC TCTTCTTACG CTTGAGAAGA GCTTCTCCGA CATGGCTTCT 120  
 TCCTTACTGA GCAAACCTT GAGCATAGAT AAGTTTGACT GGCAAGCGTG CTCTTGATTA 180  
 TTTGGCTCCC TTCCCACTAT TGTGGATAGC GAGGCGTCTT CTCATATCAG TCGTATAGCC 240  
 TATATAGTAG GATCCATCAC GACACTCCAG AACGTACATA TAAGCCTTAT GATCCATAAT 300  
 AAATCTCTTC GATTTCGGGC GTATAAGAGC CATCATCATT GTGGACAATC AAAGGAGGTA 360



210

AGACCTTAAA GCCACTTGTT GAGCCATCCT TGATCGCCTC AATCAAAAGC ATATTGGCTT	420
CCTTTTCTCT TTTTGGATAA ACAAAC TGCA GGCCTTAGG GGCTAGATTA TGTCGTTTTA	480
ACGTATCCAA AATATCCAGA AGTCGATCAG GACGATGAAC CATGGCCAAA CGCCCAT TAG	540
ACTTGAGAAT ACTCTGGGCA CTACGACAGA TTTCTTCCAA ATTAGTCGTG ATTCGTGTC	600
GAGCCAAGAG ATAATGTTCA CTCTCGTTCA GATTAGAATA AGGATTCACC TTGAAATAGG	660
GTGGATTACA CAAAATCATA TCCACCTTAC TCCCTGAAT GTGAGCAGGC ATATTTTCA	720
AATCATCGCA GATGACCTGC ATTTGCTCCT CTAATCCATT CAAACGGACA GAGCGTTCAG	780
CCATATCCGC CAAACGCTCC TGAATCTCAA CAGACAATAT CTGTGCTTGA GTACGAGTGC	840
TAGCAAAAAG CCCCCTGCT CCATTCCCAG CACAGAAATC CACAATCAAC CCCTTCTTAG	900
GAAAACGTGG AAATCGTGAT AAGAGAACAC TATCCACCGA ATAGCTAAAA ACCTCTCTAT	960
TTTGAATGAT TTTGATATCT GTCGAAAAGA GCTGGTTAAT GCGCTCTCCT GATTTTAATA	1020
ATTGTTCTTC TTCCATGGTC CTATTATAGC AAATTCATAT TAACATTACA AAAAAATAA	1080
AACTCTAAAC TACTTCTTCT TTTTAAATG GTGCAGGGCT TCTCCAGTCC AGATTGGTAG	1140
CATTCTCGA AAGGGAGCAA AGCCGTAGTT AAAGCGGTCG CTTGAAAAGC GTCTCCGTCT	1200
AGGAACTGG TACTTTTCTT CCTCCAAAGT GCGGATAGAA AGACTGGCTT TCCCTGTAAA	1260
TTTCATCTAAA TCCACTACCT GAACCTGAAC CTCTTCATCG ACTTTCAAGG TTTTCATGAAT	1320
ATTTTCAATA AATCCTGTCC GAATCTCTGA AATGTGAATC AGCCCCGTAT CACCCGTCTC	1380
TAACCAACA AAGGCACCGT AGGGCTGAAT CCCTGTAATA CGCCCCCTTA GCTTATCACC	1440
GATTTTCATC TTAGTCCTCG ATTTCAATAG TTTCAATTAC AACATCTTCA ACTGGCTTGT	1500
CCATAGCTCC TGTCTCAACA GCAGCAATGG CATCCAAGAC AGCGTAAGAT GCTTCATCAG	1560
CTAACTGACC AAAAACCGTG TGACGGCGGT CTAGGTGAGG TGTCCACCT TGATTGGCAT	1620
AGATTTCTGC AATCGGTCTT GGCCAACCAC CACGAGTAAT TTCTTTCTTA GAATAAGGTA	1680
GGTGTGGTT TTGCACGATA AAGAACTGGC TGCCGTGGT ATTTGGACCA GCATTGGCA	1740
TGGAAAGAGC ACCACGGATA TTGTAAAGCT CTTCTGAGAA TTCATCCTCA AAAGATTCCG	1800
CGTAGATTGA CTCGCCACCC ATACCAGTTC CAGTTGGGTC TCCACCTTGG ATCATAAAGT	1860
CCTTGATAAT ACGGTGGAAA ATGACACCAT CATAGTAGCC ATCTTTTGAA AGAGATACAA	1920
AGTTAGCCAC TGTTT TAGGA GCATGTTT CAG GAAAAAGCTT GATACGTAAG TCTCCGTGAT	1980
TGGTCTTAAT AGTCGCAAGA GGACCTTCTA CTGTTTCAAT GTCTACTTGT GAAAAATGCA	2040
ATTCTTTTTC TACCATACCA AATACTTCTA AGGCAGCAAA AATGCCATCT TCTTCTAATG	2100
TTTTTGTAAT ATAATCTGCT TTTTCTTGA TTTTATCATG AGAAATTCCC ATGGCAACGC	2160

TGATTCCAGC ATAATCAAAG AGTTCCAAGT CGTTGAGACC ATCTCCAAAA ACCATGACCT	2220
TCTCTGGTTT CAAGCCAAGG TGTTCACAA CCTTTCCAC CCCCGTCGCT TTGGAGCCTG	2280
AAATCGGCAC AATATCAGAC GAATGTTGAT GCCAACGAAC CATGCGAAGT TTGTCTGAGA	2340
GA CTGTCAGG CAAGTGCAAG TCATCTCCCT TATCTTCAAA AGTCCACATC TGATAGATAT	2400
CTTCTTTTTC ATGGAAATCG GGATCTACAT CTAAGTCGGG ATAAATTGGA TTGATAGCTT	2460
CACTCATCAT ATCGGTGCGA GTCGACAACT TGGCATCATG ACTCCCAACC AAGCCATACT	2520
CAATTCCTTC TTGCTTAGCC CAAGAGATAT ACTCCTCAAC ATCTGACTTT TCAATCTGAT	2580
GCTGATAAAT GACCTGACCT TTTTATCTT CGATATAAGC CCCATTCAAA GTTACAAAAA	2640
AGTCAGGCTT GAGATCACGA ATCTCTGGAA CAACACCAAA AATGCCACGT CCAGAGGCGA	2700
TTCCTGTAA AATTCCTTTT TCACGCAACT GTTTAAAAAC AGTGGGAATT GTAGTTGGAA	2760
TAAACCCTGT CTTTGAATTC CGCAATGTAT CATCAATATC AAAAAAGACA ATCTTGATCT	2820
TCTTTGCCTT GTATCTTAAT TTCGCGTCCA TCTCACTACC TCTTCAATC TAACTCTTTC	2880
CATTATATCA TAAAGTAGGC AAATCCCTA TTTTCAAAAA GTTTATCAT TTTATTTTAA	2940
TTTCTTGAT GAGAAAAGAG ACATATTTAT GAAAAAGCTC CATCGTGCTT TTAATGTGTT	3000
CTCTTGTTTT CAAACTCGTA AAAAGGGAGC CACTGATCCT AACTCGCTCT CTCATTTCAA	3060
AGCTTGTAAG AAAAGACCCG TTGGGGTCTT AATTCGCTTT CTGTGTTTCA AGTCATGAA	3120
AAAGAGACCC AACTGGGTCT TTTCTTTAAT CTTCGTTTAC GAAAGGCATC AAAGCCATTA	3180
CGCGAGCGCG TTGATAGCT GTTGTTACTT TACGTGGTT TTAGCTGAA GTTCCTGTGA	3240
CACGACGAGG AAGGATTTTC CCACGTTCTG AACGAAACG GCTAAGAAGC TCAGTATCTT	3300
TGTAATCAAC ATATTCAATT TTGTTTGCTG CGATGTAATC AACTTTTTTA CGGCGTTTGA	3360
ATCCGCCACG ACGTTGTTGA GCCATGTTTT TTCTCCTTTA TAAGTTTAGT TGTCCATTAG	3420
AATGGTAAAT CATCATCTGA AATATCCAAT GGGTTTGTTG CTCCAAATGG ATTTTCATTA	3480
CGTGAAAAGT CTGGTACTGA ATTTGTAGGT GCTGAATAGT TTGCAGTTGG TGCAGAGTAA	3540
GCTCCACCTG TGTGACCCTC ACGCACACTA CGGCTTTCCA ACATTTGGAA ATTCTCAGCC	3600
ACGACCTCTG TCACGTAGAC ACGTTGTCCT TGCTGGTTAT CGTAACTACG AGTCTGGATA	3660
CGACCTGTCA CCCCGATAAG TGAGCCTTTT TTAGCCCACT TAGCAAGATT TTCAGCCTGT	3720
TGGCGCCACA TAACGACATT GATAAAATCA GCCTCACGTT CACCATTTTG ACTCTTAAAT	3780
GTACCGTTTT CTGCAAGAGT AAAAGTCGCA ACTGCTACAT TTGATGGGGT ATAACGCAAC	3840
TCAGCGTCAC GTGTCATACG CCTACAAGT ACAACATTGT TAATCATAGT TTACCTTCTT	3900

212

ACGCGTCAAT	TTTGACGATC	ATGTGACGAA	GAATGTCAGC	GTTGATTTTT	GAAAGACGGT	3960
CAAACTCTTT	AAGAGCTGCA	TCGTCAATTG	CTTCAACGTT	AACGATGTGG	TAAAGTCCTT	4020
CACGGAAATC	TTGGATTTCG	TATGCAAGAC	GACGTTTTTC	CCAAGTTTTT	GATTCAACAA	4080
CAGTTGCACC	GTTGTCAGTC	AAAATAGAGT	CAAAACGTGC	TACCAAAGCG	TTTTTAGCTT	4140
CTTCTTCAAT	GTTTGGACGA	ATGATATAAA	GAATTTTCGT	TTTAGCCATT	GATATGTTCC	4200
TCCTTTTGGT	CTAATGACCC	CAAGACTTTG	CAAGGGGTAA	GTGAGGTTCC	CTCACAATAA	4260
ACTATTATAC	TAGAAAAAAT	TTTTTTACGC	AAGTAAAAAC	ACTAGAATTC	GAAAAAACGC	4320
CACATGGGCG	TTTTCTGTT	CTTATGGTTT	GATACGGTGC	AACATACGTG	GGAATGGAAT	4380
AGCTTCACGG	ATATGTTTTG	TTCCTGCTGC	GAAGGTTACC	ATACGTTTCA	TACCGATACC	4440
AAATCCTCCG	TGTGGAAGTG	TACCGTATTT	ACGAAGGTCA	AGGTAGAATT	CATATTCTGT	4500
ACGATCCATG	CCAAGTTCAT	CCATCTTAGC	GACAAGGGCA	TCGTAATCTT	CCTCACGCAT	4560
AGACCCACCG	ATAATTTCTC	CATAGCCTTC	TGGAGCAAGC	AAGTCTGCAC	AAAGCACGCG	4620
CTCTGGATTT	CCAGGAACTG	GTTTCATGTA	GAAGGCCCTG	ATGGCTGCTG	GATAGTTCAT	4680
GACAAATGTT	GGCACACCAA	AGTGGTTTGA	AATCCAAGTT	TCGTGTGGTG	ACCCAAAGTC	4740
ATCACCATGC	TCAAGATGCT	CGTAGTCAGC	ATCTTCATCA	TTTTTCATGCT	CTTGCAAGAG	4800
GTCAATGGCT	TGATCGTAAG	TGATACGTTT	GAATGGCTCT	GCAATGTAGC	GTTTCAAGAG	4860
TTCTGTATCA	CGTTCCAAGG	TTTCCAAGGC	TTGAGGCGCG	CGGTCAAGAA	CACCTTGTAG	4920
AAGAGCTTTC	ACATAAGCTT	CTTGCAAGTC	AAGCGACTCA	TCATGTGTCA	AGTATGAGTA	4980
CTCAGCATCC	ATCATCCAGA	ACTCAGTCAA	GTGACGGCGT	GTTTTTGATT	TTTCAGCAGC	5040
GAAAACTGGA	CCAAAGTCAA	AGACACGACC	AAGAGCCATA	GCCCCTGCTT	CTAGGTAAG	5100
CTGACCTGAT	TGGCTCAAGT	AGGCTGGCGT	TCCGAAGTAG	TCAGTTTCAA	AGAGTTCTGT	5160
AGAATCTTCT	GCCGCATTTT	CTGAAAGAAT	TGGGCTGTCA	AACTTCATAA	AACCGTTCTT	5220
GTCAAAGAAC	TCATAAGTTG	CATAGATAAT	AGCGTTACGG	ATTTGCAACA	CAGCTACTTG	5280
CTTACGAGAG	CGTAGCCACA	AGTGACGGTT	ATCCATCAAA	AAGTCTGTTT	CGTGTTCTTT	5340
TGGTGTGATT	GGGTAGTCTT	GAGATTCACC	GATCACTTCG	ATGTCTGTGA	TGTCCAACTC	5400
ATAGCCAAAT	TTAGAACGTT	CGTCTCTTTT	GACAATACCT	GTACATAAAA	CAGACGTTTT	5460
TTGGCTCAAG	CGTTTGATAA	CATCAAACCT	CTCAAGTCCC	ACTTCTTCAC	CAAATTTTTT	5520
GACAAAGTTT	GGTTTAAAAG	CCACACCTTG	AAAGAAGGCT	GTTCATCAC	GCAATTGTAA	5580
GAAAGCGATT	TTTCCTTTTC	CTGATTTGTT	GGCAACCCAA	GCGCCAATCG	TCACTTCCTG	5640
ACCAACATAG	TCTTTTACGT	CAATAATCGT	TACACGTTTT	GTCAATTATT	TTCCTTTTCT	5700

213

TTTTTATTCT TTATGGCAAA CCACCTCTAT ATTGTTCCCA TCCAGGTCAA TCATAAAAGC	5760
AGCATAGTAA ATCGGATGCT CACTTCGATA ACCAGGAGCC CCATTGTCTC GCCCACCTGC	5820
CTCTAAGCCA GCCTCATAAC AAGCCTGAAC TTCTTCCTTA TTTTCTGCTA AAAAAGCAAA	5980
ATGAACAGGA TCTTGTGTTC CCTGAGTCAG CCAAAAATCA CCACCAGGAT GAGGGCTGTT	5940
CGGGGATAGA AAACATAATTA GAGAACTAGT CTTAAAAGCC AATTTATAGT CCAAAGGAGC	6000
GAGAAACTC CTATAAAATC CTTATGAAAT TTGTAATCC TTTACCTTAA TCTCAAATG	6060
ATCAATCATT CTCACTACCC ATAAATGCTT TCAAGCGTTC GACTGCTTCT TTAAGCGTGT	6120
CTAGGTCTGT CGCATAGCTG AGGCGGACAT TTTCTGGTGC TCCAAATCCA GCTCCTGTTA	6180
CCAAGGCCAC TTCGGCTTCT TCTAAGATAA CAGTTGTAAA GTCTGTCACA TCCGTGTAGC	6240
CTTTCATCTC CATGGCCTTT TTGACATTTG GGAAGAGATA GAAGGCCCTT TCGGTTTGA	6300
CCACTTCAAA TCCTGGTACC TCTGCAAGGA GGGGATAGAT GGTATTAAGA CGTTCCTCAA	6360
AGGCCTGACG CATGCTTTCT ACAGTATCTT GCTCACCTGA TAGAGCCTCA ACTGCTGCAT	6420
ATTGGGCTAC TGCTGACGGA TTCGAAGTTG TTTGACCTGC AATCTTGGAC ATGGCAGCGA	6480
TAATGTCTGC TTCTCCAACG GCATAACCAA TCCGCCAACC AGTCATGGCA TAAGTTTATG	6540
ACACACCATT GATGACCACT GTTTGCTTGC GAATCGCTTC CGATAGGCTA GAAATCGGTG	6600
TGAACTCATG ACCATTATAA ACCAAGCGGC CATAGATATC GTCTGCTAGG ATGAGAATAT	6660
CATTTTCTAC AGCCCAGTTT CCAATTGCCA AGAGTTCCTC ACGGGTGTAA ATCATACCTG	6720
TGGGATTAGA TGGCGAATTC AGCACCAAAA CCTTGGTCTT GTCAGTGGCA GCTGCTTCTA	6780
ACTGCTCTAC GGTCACTTA AAGTGATTGT CTTCCTTAGC AGAAACAAAG ACGGGAACGC	6840
CTTCTGCCAT CTTGACCTGA TCTCCATAGC TAACCCAGTA TGGGGTTGGG ATGATGACTT	6900
CATCACCTGG ATTGACCACA GCCATAAAGA AGGTATAGAG AGAATATTTG GCTCCCGCAG	6960
CGACTGTCAC TTGATTTGAC GCTACAGAAT AGCCGTAAAA GCGCTCAAAG TAGCTATTGA	7020
CCGCCGCCTT AAGCTCTGGC AGACCTGAGG TTAAGTATA AAAAGAAGCA CGCCCATCTC	7080
GAATCGATGC AATGGCGGCA TCTTGATAT TTTTGGGAGT AGTGAAATCT GGCTCACCCA	7140
AGGTTAGAGA CAAAATATCT CTACCCTCAG CCTTCAGTGC TTTGGCACGG GCTCCAGCAG	7200
CCAAAGTCAC ACTTTCTTCC ATTTCTAAAA CACGGTTGGA TAGTTTCATA GGCCCTCCTT	7260
GTTGACCAAT GCTCCTGTTT CAAAATCTAC TAGATAAAAA TCAGATCCTG ACTTAACTTC	7320
CCAGATTGGC TTATCTTGAT AACGGCCAAA GGTATCTTG TCAATCTCGC CAGCTCCCTT	7380
TTCTTAGAA ACCGTTTCTG CTTTTTCTTG TGAAACACCC TGATTTAGCT GATAAACGTA	7440

214

AATCTTATCG TCATCTTTAC CAATCAGGAC AGCAAGCGCT TCTTGCTGTT TGTACGACC	7500
AAGAACGCTG TAATAAGATT CCAAGCCATT GTATAAATCA ACCTGATCAG CCTGCTCTAA	7560
TCCTGCATAC TGCTGAGCTA ATTTTCTCC TTCACTTTTA GCTGTTTGAT AGGGTTTCAT	7620
GCTAAGAGAA ACCATATACA GAAAGGAACC ACTGATAACC ACAAACAAA TCCTCATCCC	7680
TAGACCATAC TGCCACAGTA GATTATTTTT TGCTTTGTTT TGTCTTTTTT TCACTCGTCT	7740
ATTTTACCAT CTATTAAGCT TTATTACAAG TGAATATAAG AATACTCTTC GAAAATCTCT	7800
TCAAACCACG TCAGCTTTAT CTGCAGACCT CAAAGCTGTG CTTTGAGCAA CCAATTCTAT	7860
TTCTCCCTTC AAACAAAACC GATTTTGAAA GTGAAACAGT TCTTACTTTT TCAGTCACAA	7920
ATGATTAGAG TTTGCCGGG	7939

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9897 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CCGCTCTACC GTCAAATAAT TACCATTTTG TTAAATACCG AAATTTTTAT CTAAGAAAA	60
TTCAAGTTGGT CTGTTGGTAC GATCGTCGTA TACAGTACCA TTCTCAGCAA TAGTATAATT	120
GTAATCAGTA TCACCTTGTT TCCTTAATTT AAGGTAATAA TTACCATCAA TTTGTTTATA	180
ACCTGAATCT TTTCTAGTTG CTCTCTAAA ACTTACTCCA GCAGGCATCA CATCAGCAAA	240
CATGAGTACT TGTTTGTCT TTTTTCAAC AATAACAGAG TCAATATAGG TTGCACCACC	300
GCTGATTTGT AAGTCACGTC CACCAACTTC ACGAGGCCAT TCTAATGGTA CTGGCGCAAA	360
ATCATCGAAT GCCAATGTTA ATTTTGTTTT AGTCCATGTC TTACCATTAT CATCACTATA	420
ACTTGTAGCA ATATTAATTT TATTCAAGAA ATCATGAGTT CCACCGTAAC GAGCGTCAAT	480
GCTTGAAAAT ACCCGACCAT TGCTAAAAGT ATACAGAAGT GGAATACGGA AATAGTTAGA	540
ACCTGTTGTA TCATTAGCCG TATAAATTAA ATGTCCAGTA ACAGCGTTTG TTGTCATCTT	600
TTTAACAGTT TCTTCATCCA ATGCACTATT AAAGAATTTG ATATTTTCTA GTGTTCCGTT	660
AAAACCAAAC GCCGTTTTTC CTGCACGTTT CACTCCCCCA AGCATATAGT AATCAATACC	720
TTTAATATCC TTGATGTTTA GGAAATTATC CACTTTCTTT TCTACTACTT TTGTACCATT	780
TGCGTATAAA GAATATGTTT TTTTGACTGA ATCTGCTACT ACTGCAACAG TGTTAGTCAC	840
AGCCTCTTGT TTGTACTTAC CCCAACTGA AGCAGGTCTG GATACTAGGT TATTTTTATT	900

215

GGAAGAAGTA TCACGCGCTT CCATCCCCAA CTCACCATTG TCTCTAAGGA ACACATCTAC	960
ATAACTATTT TGTGACCGG GTTTGGAATT AGATAATCCA AACAGAGCTT GTAAGCCTTT	1020
CTCACTTGAC TGATTGTACT TAATCACTAC AGTAAAGTCA CCGCTAGTAA ATTTATCCTT	1080
TAACTCTTTA GTAACATTTT CTCCGCCCCC TGTAAAGTA ACATTATTTT TTTCTAAGAC	1140
AGGAGTTTCT TCCGCTGTAG AAGATGGATC CTTAACAGTA GTTTCAACTG TTCGAGGTTG	1200
TACAGTAACT TCCGAAGAGT TATCCGATGT AGGTTGTACT TCCGAAATCG GAGTCGTTGG	1260
TGCAACAGGT TGCACCAACT TTGGTGTGCA TACTTCAGAA GTTTCAGTCT CCTGAGCTGC	1320
AACTGAGTTA GCAACAAATG CTGATAATAC CACTACAGTA CCTAAGGTTA CATATTGTTT	1380
AATATTTTTT TTCATTTTAT TTTTCCTCGT TTAACCTTT GATAACAAGT TTTTAAACAG	1440
TTTCATCATT GCAATGAATC TTTGGTTGGT GAAGATCTTC TTCAAAAGTC ACCAACATAT	1500
TCCCTGGAAG CAATTCAACA ATTTGATAGT CTTTGCTATC GTAAAAAGCA ATATCCTTCT	1560
CTTCGCTAAA AGGTACACGT GACTGGGCAC GAACTGGGGA AGTTACTGCC ATTTTTCAG	1620
TATTTTCAAC AACAATATGA ATATCTAAAT ATTTCTTATG AGTTTCAAAA ATATCTCCTG	1680
GAACTCCATC AGCTAGATAA GTCATACAAT TTGCAAAAAC ATTTTCCCCG TCAATATCAA	1740
TTTTTCCATC AACTAAATCT GTCAAATTTG TATTTTCTAA AAAATCACAG ACTTTTGAAA	1800
AATATTTATT GACAGAAGCA TATCGTTTAA AATCAGATTG TTCAGAAATA ATCATATTAT	1860
TTTCTCTTTT CTATTAGTGA CGAACTTCCC AACTTGAATC CGCTTTAATT TCTGTAATAT	1920
CATGAATCGT TGTATATTTA GGTGCAGATA CTTTATTTCC AGTAAGAACA GATACAATAT	1980
AACCTGAAAC TACTGATACA GAGATTGAAA TCAATGAATA TGCCCAGTAG CTAACAGCTG	2040
TTGGAGGAAG GAAGTATTTA ATAAATACCA TGACGATGGT TGATACAATC AGCGCTGCAT	2100
AAGCACCTTG TTTATTTGCT TTTTLAGAAA CAAATCCAAG AATAAATACA CCACCAAGTA	2160
GACCAAGTAC AAGTCCCATG AACTATTGA ACCATTGCTA TGCAGATTTA ATATCTGAGT	2220
GAGCCATGAC AATGGAAACA CCAATTGAGA ATAAACCTAC TGCTAGAGAT ACGAATTGTG	2280
CAATTTTCGT ACGACGATTG TCTGACATAT TTTTAGAAAT GACATCTTGA ATATCCAATG	2340
TCCATGAAGT TGCAACAGAG TTCAAACCTG TTGAAATAGT TGATTGAGAT GCTGCATAAA	2400
TCGCTGCCAA GATCAAACCT GTGATACCTA CTGGTAACTG GTATGCAATA AAGTACATAA	2460
AGATTTGGTC TTGAGGGATA TTGCTAGCTG CACTATCTGC ATTTTGTACT TGATAGAATA	2520
CGTACAAGCC TGTACCAATC AAGTAAAAGA CTGTTGCAGT TGCAAGTGAC AAAACACCGT	2580
TTGTGAACAA CATCTTATTA AGTTTCTTAA TATTTTGTGT TGTAGTAAAA CGTTGAACCA	2640

216

AATCTTGAGA	TGAAGCATAG	GAAGACAAGA	TTGTAAAGCC	TGAACCCATC	ACAATTAAAA	2700
AGATGGAGTT	TGAAAGCAAG	TTAGGATCGA	AAAGTTTTTC	ATTTGCAGCA	AGGAATTTCC	2760
CGTTTGCTAA	TGTTTCTGCT	ACTGCACCAA	AGCCACCTTT	AATATTAGCA	ATCAGTACAA	2820
ATAAAGCTAA	AACGACACCA	CTAATCAGAA	TCACACCTTG	AATAAAGTCT	GTCCATAATA	2880
CGGATTTTAG	ACCACCAGTA	TAAGAATAAA	CAATTGCAAC	TACACCCATC	AAAATAATCA	2940
AAATATTGAT	GTCAATTCCT	GTCAATACTG	ATAAACCAGC	TGATGGGAGG	TACATAATGA	3000
TAGACATACG	TCCCAATTGA	TAAATAATAA	ACAAGAGTGC	TGAAATAATA	CGAAGTGCTT	3060
TAGAATTAAA	ACGTTTATCC	AAGTAATCAT	ATGCCGTATC	GATGTCTATC	CGTGCAAAGA	3120
TAGGTAAGAT	AAAACGAATT	GTCAGTGGAA	TAGCTACTAC	CATCCCTAAT	TGAGCAAACC	3180
ATAAAATCCA	GCTACCTGCA	TAAGAGCTAC	CAGCGAGTCC	CAAGAAGGAA	ATCGGACTGA	3240
GCATTGTGGC	AAAAATGGAT	ACCGAAGTAA	CATACCAAGG	AACCGAACCA	TCTCCTTTAA	3300
AGAACTCTTT	TCCTTTTCATC	TCTTTTTTAG	AGAAATAGAT	ACCTGCAACC	AACACCGCAA	3360
GTAAATAAAC	AATCAAGATA	ATTAAGTCAA	TTATTGTAAA	TCCTGTTGTG	CCCATAACAT	3420
ATCTCCATAT	TGATTTTATT	TATTATAAAA	ATTCTTTTCG	TGCTTGTGTA	ATAAGTTCTG	3480
CTGCTTGTTT	TGCAACTTCC	AAGTCACCTT	CTGCCAATGC	TTCTAAAGGT	TGACGAACAG	3540
AACCTAAATC	AAGTTTTTCA	TTTAGACGCA	AAACTTCTTT	TGCTACAGCA	TACATATTTG	3600
CCTTACCTGA	TATCATCTTA	TAGATAACTT	CATTGATAGC	ATATTGAAGT	TTTTTAGCTG	3660
TATCTAAATC	TCGTTCTTGA	ATCAAACTTT	CCAATTTCAA	GAACAAATCT	GGCATAACGC	3720
CATAAGTACC	ACCAATACCA	GCTTCTGCTC	CCATCAAGCG	ACCACCAAGA	TATTGTTTCAT	3780
CTGGACCATT	GAATACAATG	TAATCTTCTC	CACCTGCAGC	TACAAACATT	TGAATATCTT	3840
GTACAGGCAT	AGAAGAATTT	TTAACTCCAA	TCACACGAGG	ATTTTGACGC	ATTGTTGCAT	3900
ACAAACTACC	AGTCAACGCA	ACCCCTGCCA	ATTGTGGAAT	ATTATAGATA	ATAAAATCTG	3960
TATTTGACGC	AGCTTCACTC	ATTGCATTCC	AATATGCTGC	GATTGAATAC	TCTGGCAATT	4020
TGAAATAAAT	AGGTGGGATA	GCTGCAATAG	CATCGACTCC	AACACTTTCT	GAATGTTTTG	4080
CCAATTCGAT	ACTATCTTTC	GTGTTATTAC	ATGCAATATG	GTTGATAACT	GTTAATTTAC	4140
CTTTAGCAAC	TTCCATAACA	GCTTCAATAA	TTTGTTTACG	ATCTTCTACA	CTTTGGTAAA	4200
TACATTCACC	TGAAGAACCA	TTTACATAGA	TACCTTTTAC	ACCTTTGTCA	ATGAAATATT	4260
GTACCAGAGA	TTTTACACGA	TCTTGGCTAA	TTTCACCATT	TTTCATCATG	CAAGCATAAA	4320
ATGCAGGGAT	AACGCCTTTG	TATTTAGTTA	AATCTTTCAT	CAGATTTCTC	CTTTATATTG	4380
TTTTTTATTT	GATGACATTA	ATAAATCGCT	GAGCAATTC	TTTTGGACGT	GTAATCGCTC	4440

217

CACCAATGAC TACACTGGTA ACACCTAAAC TATAAGCTTT TTTTAATTGT TCTGGATAAT	4500
GAATTTTTC TCGGCAATTA CCGGAATATT AAAATCAGCC AATTTTTC TTAGTTCAAA	4560
ATCAGGCTCA TCTGATTGTA CACTTGTA TGTGTAACCT GATAATGTTG TACCAACAAA	4620
ATCAACGCCT GATTTAAATG CATAGAGACC TTCATCTAAA TTAATTACAT CCGCCATCAG	4680
CAATTGATTC GGATATTTTT CTTTTATTTT TTTGATAAAT TCACTGACAA CTAAGCCATC	4740
ATATCTTGGT CTAAAGTTG CATCAAATGC AATGACTGTT GTTCCGCATT CTACAAGTTC	4800
ATCTACTTCT TTCATCGTAG CAGTAATATA TGGTCTTGA GGTGGATAAT CCCTTTTGAT	4860
AATTCCAATT ATTGGTAAAT CTACTACTTT CTGAATTGCT TTAATATCAC GCACAGAATT	4920
TGCGCGAATG CCCACTGCTC CTGCCTCTAA AGCTGCTTTA GCCATAAAG GCATCAAGCT	4980
AAATTCTTCA TTATAAAGG CTTCACCAGG TAAAGCTTGA CAAGAAACAA TGAATCCACC	5040
TTGAACCTGG CTATAAAAT TTTCTTAGT CCAAATTTGG CTCATTTTAT TATTCCTCCT	5100
TATGGATAAT AGTTTGATTG TAATAATATT GTCTCTCTGG ACTTTCAGTA TAATTAGAGA	5160
ATAAGCAGTC TGTAATTAAA AGTATTGGAA ACTGAGGTGA TATGCGATTG CCATACGAGA	5220
GATGATCGGT CGAAGCTAAT AACAATAGTT CATCAAAGAA ACAATCTTCT TCGTCAAATT	5280
TTCTTGTAAT CATTAAAACT GTTTTAGCGC CTTTATCTGC AGCTTTTTGT AGACCTTCTA	5340
GTACAATATC AGTTTGACCT GAAATGGATG CTCCAATGAC AAGGCAATTT TCATTAAGTA	5400
GTAAGCTACT CCACAAAATC ATATCCTCGT CTGATAATAC TTACCAATC ACTCCGAGAC	5460
GCATAAATCT CATCTTCATT TCTTGTAAG CAAGAACAGA ACTTCCTTTA CCGTAGAGAT	5520
ATACACGCTC AGCAGTTTCT ATCATCTCAG CAATACGCTC AAGTTGAAC TCATCAAGAA	5580
CCGTGTAAGT TTTTCTCAAC ATTTCTCAT AGTCGGATAA AACTTTTTCT GTTGCTCTG	5640
TATATAATGC CAACTTTTCT TTCTCATGAA TCATCTCTTG GTATTTGAAA ATGAATTGTC	5700
TAAAACCTTT AAAACCACAT TTTTTCGCA ATCCAGTCAA TGTTGCTTTG GATACATTAA	5760
GCTATTCGCA CAATGCTTTA GATGAATAAT CATTCAAGAG TTGCTGTTTT AAGAAGAATT	5820
TAGCAATGTC TTTTTCAGCA TATGCCATAT TTGGTAAGTT AGCTTCTATC ATTGGAATTA	5880
GTCTTTTTG CAGTAACATA TGAGCTCCTT AGTTGAAGTA AACGTTTACA TTCTTTATTT	5940
TAACACTTTT TTTTTTTTTC AATATTTTTC ATAAATTAGA AACTAGTTTC CAATTTCTTT	6000
CGTTTCATAA CAGAACAACA AACATAAAAA TATAATAGTT TTTATCTTTT TTATCGTAAT	6060
TATATGTATT GTAAGAACGT TTATCACTAA TAATATGTTT ATATTAAAA ATTTTAGTAA	6120
TATTTTATTT TGGTTTTATT ATTTCTTTTC GGAATTTCTA TATAATATTT TATTTCTAAA	6180



218

AAAATTGAAA AAATATTTCT AGTTTCTTTA TTTTATATAG GTAAATATATT TTATTTCTAA	5240
ATTAAGAG AATCCCATAA AAACACAGA TTTATGAGAT AAATCAGGTC ACCTATTTTA	6300
AAAAAGCAGC AAACATATAA CTAAGAGTT CCACACCAA TGTAACCCCA TACTTCCCA	6360
TAAATCAGAT TTATAGCGCA CCATACCTAA AAACATTCCA AGTGAAACGT ACAGACACCA	6420
AGCTAGAATG GTTCCTGGAT GATGTACTAA GGCAATATA ACACTTGTC AAGCAACTCG	6480
AATATCTAAT TTTCTAACCA AGTTCCATAA AATTTACGA TACAGAAATT CTTCAACCAT	6540
ACTCGCATTC ATTAAGAACA ATAAAAATGA AAACCAAGGA ACTTGATGTT GAAGGCCAAT	6600
TAAATTTGTT TGATTCGTGC TTCCTTGAGC ATGAATCAGG CTAAACATA GACTTATAAT	6660
CAGTAGACTA GCTAGTCCAA TACCAAGGCA TTTATCCTA GTTTTCATAT TGACCTTGAC	6720
CACTTGTTTT CGTTGACCAT ACATCCATAA AAAAGAAAA AGAGACGCAC CATAGAGAAC	6780
CTGTAGTATA GTTAACTCAC CGATACAAAG AAATTTCAAT AAGTATAGAG ATACCAATAG	6840
GACATTTACT TGTTGGAATA TATAAAGTG AATTATTCTT TTCATAGTTA CCTCCGAAAT	6900
AAATCTTCAT AATCTAAATC TAATATCTGC ACAATCCTTT CTACCCATGG ACTTTGAGGC	6960
ATTCGTTGTT CCATCTTGTA GTGGCGAATC TTTTGATATA AACGATTCAA TTCACTTGGA	7020
TAGTGAAACT CTCCCGCAA CATTTTCTG GTTAACTCAA TCCAGCTGAT ATTTCTTTCA	7080
GCCAAATAA TGGACAAGTT CTCCCAAAAT CGTTCAGCCA TATTCTTCT CTTTAGTTA	7140
GATAAATAAT GTGTTGyGC CATGTAAATC AATTGTTTCG TATCTCTGG CAATAGAGCT	7200
CTAGCCTCTT CCAAAATCAG ACTTGATAA ACCCGCTTAT TTGAAACCAC AAAAGGAAGT	7260
CCGATGGTTA GTTCAGGATT TTTTAAATTT ATCTCAACGA AATCCGTCAA TCTTAGATTG	7320
TCACGGTTCT TAAATCGTAA TAAATGGGA GATAAAACT CAAACAATC TGAAGAATAG	7380
CTCATCATCT CAATTAATTT GTCCTTGTC ATTTAGAAA CTGAATGACA AGATACCTCA	7440
ATGCCATAGT TTTGGAAGAA GTCTAAAGA AGTTGATTTT TTTGGCTATT TTTACTTAGA	7500
TAGAGATCAA TCATGGGAGA CCTCCAACAA ATTTGCTTCC ATTTGATATT CTGAGACGAT	7560
TAAGGAATCT AACAACTTTG AGAAGTTAAT CGATTCTTG TCTTCATCAT AAGCTTTTAC	7620
AGTTACTTGG GTTGTAAGTA TCCCTCTTT TCCCTCGGCT CGATAGTCTT GTCAATATAA	7680
AACAAAAACA AGATTCTGAT TATCATCTAC AAAGGCATTA ACTCGTTCTT TTATATCCTG	7740
ACTTTCAAGG AATTCATAA CGTTTGAAG ATAGGATTCA TAAATAGTG GGTAAATTAG	7800
TTTTTTATGG TAATCATCTA AAAATGTTAC CTCAAACCTA CATGGATAAT TGGGCATCAA	7860
AAATATTTGT TCATCCAGCT GTTTGATTTT TGCATCATGT AATTCTGTTT CTAATTCATC	7920
ACAATCTAGT ATTGATTCTT TATTTAATGC TTTTATCTTT TTCCTCTATT TCTTTAATT	7980

219

TCTTTGCGAT TCGGCAATC ACAGGAACGG TTACACTATT ACCAACTTGT TTATAGAGCT	8040
GACTATTAAT AGAGACTTTT CTAGCAGCTT CAAAAGCCTA ATCAGGAAAG CCATGCAATC	8100
GAAAACACTC TTTAGGAGTG ATTGTCGTA TTCTCAAACG GTAAAATTGT CCATCTATTA	8160
AAACACCAGC TACTTGGTAA ACTTGTTTAT CTCTCCTTC ATAGCTAGCC ACTACTACTC	8220
CCATTTGACC ACTAGTTGTT AACGTATTAG CTATACCTTT TCCAACTCTA CCACGACGAT	8280
ACTGAGAACT TGGTCTTTCT AAATTGATTG AATCCCCAAT CTCTGCTTGA GCATATCCTT	8340
TTTTCGTTGC TTCCCGTACT TTTAGAAATT GGATTGGTTC TGGAAATTAGT ATTTTGGGGA	8400
TTTATCTCC TCCTTGCAATC GTAGTCAGTG TTGGAGATAA GCCCTCACTT CCATAGACAC	8460
GACCTGTCTC CTAAAGCTA GTCGGTAAAT CTCCAACAAC GACAATGCCA TAACGATCCT	8520
GAGTATTTAA AGTAAACATC GGCTCTTGAT TTTCCTTAAA GCGTCTCCCA TTTTGTCTCT	8580
TGCTAATCT ATCTGGTGC ATACAAGGAA TCGCAACTTT AAATCCTTCT CCTTTACCAC	8640
GAACAAAGGT TGGCGCAAGA CCTTCTGAAT AATAGACTTT ACCGCTCATT CCACCTCTTG	8700
ATGGATTCAA ATTTCTAGT GCTTCAAAG TCTCAGACTT AGTTGCTTGA CCTTCTCGTC	8760
TGAAAGGAAA TAAGAGTCTG GTACCTTTCT TTCTAGAATG TCCGATAATA AACACCTCT	8820
CTCTGTTTTT GGGAAACGCCA AAATCCTTAC TGTTAAGCAC CTGCCACTCA ACATCAAACC	8880
CCAACTCATC AAGTGTGGTA AGTATTGTGG TGAACGTCCG TCCCTTATCG TGATTGAGTA	8940
GGCCTTTAAC ATTTTCAAGA AAAAGAAAAC GTGGTTGGAT TTGTTTGGCC GCCCGAGCAA	9000
TTTCAAAGAA CAAAGTTCTT CTAGTATCTT CAAATCCCAA TCGTCTTCTT GCGATTGAAA	9060
ATGCTTGACA AGGGAATCCC CCACAGATGA CATCGACTTT CCCTCTAAGT TTTTAAATT	9120
CGTCATCTGA AACATCTCGT ATGTCATGAA ATTCTATTTT TCCTTCCGTT TGAAAAATGG	9180
ACTTATAAGA TTTCTAGCA AATTTATCAA TCTCACAATA TCCCAAGCAC TCATGCCCTT	9240
GAGCTTCCAT TCCCATCCTA AAGCCTCCTA TCCCAGCAAA TAAATCTAAA ACCCAAATCA	9300
TCATACCTC TCTCAACTAG ATGTAACCTA CAAAACCCCT GACCTCATGA GCCACTTCT	9360
TCCTCTCAT GAGGTCAGTT TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT	9420
TTCTCAAAA GGGCAGACTC CTCCCTTGGT TCGTCACAG ATTTTTTCAT CTCGACTGTT	9480
CTTTAATGCA TCATTAAACGA CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC	9540
AGGTTGACTT TTCTAATCCT AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA	9600
CTAGACAATT TGAGGAGCTG CTTGCGTCTT GTTCGAACAC ATTTTCTTAC CACGTGAAGA	9660
AAAAGATGGC GGAAGCGTTT GATTGTAAA GTTTGGAAGT CACCTCCAGC TAGATGTTTG	9720

220

AGAAAAAGAT AGAGATTGTA GGCGATACAG CTCATCATCA TACGAACTCG TTTTGTGATTA 9780  
 AGGTTGAACT ATCCGTTTTA TCGCCAAAAA ATCCCTCCTT CATCTCCTTG ATGAAATTCT 9840  
 CGGCTTGACC ACGTCCACGA TAAAGCTGAA ACTGGTCTTG GCTTGTTCGG GTACCGA 9897

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8148 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CCGTGGAACA AGCCAAGACC AGTTTCAGCT TTATCGTGGA CGTGGTCAAG CCGAGAATTT 60  
 CATCAAGGAG ATGAAGGAGG GATTTTTTGG CGATAAACG GATAGTTCAA CCTTAATCAA 120  
 AAACGAAGTT CGTATGATGA TGAGCTGTAT CGCCTACAA CTCTATCTTT TTCTCAAACA 180  
 TCTAGCTGGA GGTGACTTCC AAACCTTAAAC AATCAAACGC TTCCGCCATC TTTTCTTCA 240  
 CGTGGTAGGA AAATGTGTTT GAACAGGACG CAAGCAGCTC CTCAAATTGT CTAGTCTCTA 300  
 TGCCTATTCC GAATTGTTTT CAGCACTTTA TTCTAGGATT AGAAAAGTCA ACCTGAATCT 360  
 TCCTGTTCCT TATGAACCAC CTAGAAGAAA AGCGTCGTTA ATGATGCATT AAAGAACAGT 420  
 CGAGATGAAA AAATCGTGTG ACGAACCAAG GGAGGAGTCT GCCCTTTTGA GGAAATCTAG 480  
 CGAGGAAAAA CGATACTGGA ACAGCAGAAA GTAAACTGA CCTCATGAGG AGGAAGAAAAG 540  
 TGGCTCATGA GGTGAGGGT TTTGTAAGTT ACATCTAGTT GAGAGAGGTA TGAATGATTT 600  
 GGGTAAATAC AATGAGCTTG AAAGAAGTAG CAACTCACC AAGCGCCAAT TCTTTGAGAA 660  
 TCAGATGCTG GATTATACCA TCATTGCGCA TGAGAGTTTT GAAATCATCC GTCATTCTGT 720  
 CTACCAGACA GATGATCGTG AAGTGAAAA TGCTCTGGCT TTTGAAGTGA AAAATGATGA 780  
 AACAGACAAG CTGATTCTGT TATTAAGCGA GGATATTGGT GTAGGTGAAA AATTGTGCCT 840  
 CGTTGACGGA ACAAATATGC GTGGAAAAATG TTTAGTATAT GATAAAATAA ATGAGAGAAT 900  
 GATTCGCTTG CAGTGCTAGA AATAGGCATT TTGAATAGTG AATATGTTAT AATAAGTATT 960  
 AGTAGGAGGT GTTTTAGATT GGAGAAGAAA CTGACCATAA AAGACATTGC GGAAATGGCT 1020  
 CAGACCTCGA AAACAACCGT GTCATTTTAC CTAACGGGA AATATGAAAA AATGTCCCAA 1080  
 GAGACACGTG AAAAGATTGA AAAAGTTATT CATGAAACAA ATTACAAACC GAGCATTGTT 1140  
 GCGCGTAGCT TAACTCCAA ACGAACAAAA TTAATCGGTG TTTTGATTGG TGATATTACC 1200  
 AACAGTTTCT CAAACCAAAT TGTTAAGGGA ATTGAGGATA TCGCCAGCCA GAATGGCTAC 1260

221

CAGGTAATGA TAGGAAATAG TAATTACAGC CAAGAGAGTG AGGACCGGTA TATTGAAAGC	1320
ATGCTTCTCT TGGGAGTAGA CGGCTTTATT ATTCAGCCGA CCTCTAATT CCGAAAATAT	1380
TCTCGTATCA TCGATGAGAA AAAGAAGAAA ATGGTCTTTT TTGATAGTCA GCTCTATGAA	1440
CACCGGACTA GCTGGGTAA AACCAATAAC TATGATGCCG TTTATGACAT GACCCAGTCC	1500
TGTATCGAAA AAGGTTATGA ACATTTTCTC TTGATTACAG CGGATACGAG TCGTTTGAGT	1560
ACTCGGATTG AGCGGGCAAG TGGTTTTGTG GATGCTTTAA CAGATGCTAA TATGCGTCAC	1620
GCCAGTCTAA CCATTGAAGA TAAGCATACG AATTTGGAAC AAATTAAGGA ATTTTACAA	1680
AAAGAAATCG ATCCCGATGA AAAAATCTG GTATTTATCC CTAAGTTG GGCCTACCT	1740
CTAGTCTTTA CCGTTATCAA AGAGTTGAAT TATAACTTGC CACAAGTTGG GTTGATTGGT	1800
TTTGACAATA CGGAGTGGAC TTGCTTTCT TCTCCAAGTG TTTGACGCT GGTTCAGCCC	1860
TCCTTTGAGG AAGGACAACA GGCTACAAAG ATTTTGATTG ACCAGATTGA AGGTCGCAAT	1920
CAAGAAGAAA GGCAACAAGT CTTGGATTGT AGTGTGAATT GGAAAGAGTC GACTTTCTAA	1980
AATGAAGGAA AATGACTTGC AATCTCTGTT AAGAAATAAA ATAATCCAC CTAGAACAAG	2040
CTAGTGCGGA TTATTTGCCT ATGAAATGAG AAATTATGGG AGCAAGCTCC TAAATCAACT	2100
GTTTTTGATC TACTTCTTA ACTACTTGAT AAAAGTTATA GAAGTAGGCC AAACTTGAAA	2160
TGATGGTTAC GACTAGGAAT ATTGAAAATT TCCATTGGAC AGGGTTGGTT AAAAGTTGTG	2220
GAAAGGATAT GAGGAGAAAAG AAGAGGGCTG CGTTGAGGAC AGGTATCCGT TTTGATTGTA	2280
TTTTCTCAAG TCCTTTATTG AGCGCAGGAA GAAAGAGGAG TAGGAGTAGT AAACTGTAT	2340
GAGAAATAGC TCCTGAAGTA AGGCCGAAGA AAAGGAAAAT ACTGATAAAA ACATGAATGA	2400
TCAGTAGTCT AGCTAGTGAT TTCATAAGGC ACCTCCTAAT CCTGGTCTTT TTTAGCTCTT	2460
GCAATACGAA GTGAGTCGAC AATATGTATC ATCACTCCGA AAAAGAAAGC TCCCAGTATA	2520
GTTTTAAAAA TATGTTTTGT ATTTAGAAGA GAACTGATAA AATTTGGATT TTCACTTGT	2580
AGGGTATCAA TGAGTGGAAT TATAAAAAAT ATCACTGTTT CATAAATCGA ACCTGCTTTC	2640
AGACCAGGAT AACGTAAGT TTTCTTTCT TTTTCATGA GTTTCCTCCT AATCCTCATC	2700
TTGATTTTTC TTAGTTTTTG CAATGCGACG GGAGATGAGG AACTGTATGC TCGCTCCGAA	2760
GAAAATAGAA CCGAGAATAC TTGATACACC ATTTCTTATA GTGAGAAGAG AATGAAAATA	2820
GTCCTGACCT TCATCTATGA GTATCCTGAG AAGAGGAGTT ATAAAAACA TCCATAGACC	2880
AAAGAACAAA CCTGCTTCA GACCTGGGTA GTGTAGTTGC TTGCTTTCTT TCTCATTCAG	2940
CATATCTGGT TCAATGACTG TGATGCCTGT TTTTTCATT TGGTAGGTGA CATAGCCAGA	3000

222

AGCGATGAGG GCAATCACTA AAATCAGAGG AGGATAGATT AGAGCCACTT CTTGAGGGTA	3060
TTTATAGGCC AGAAGGAGTG GAATAAGATT TCCGAAAATC ATCAGATAAA AGAGGATGAT	3120
AAAGACTTGG TTCCCAATAC TATCGGCCTC ACGCCGTTTG TATTCGTCAA GGGGACCAGA	3180
AATACCGTAT GTGCGTTTGA TCAGTTTTTC AGTGAAGGTT TCTTTTTTCA TGAGTTTGCT	3240
CCTTTTTTAA AAATCTTCCT CCCAAAAGAG ACTGTTGAGG TCAGTTTGGA GGCTGCGGGC	3300
GAGATTGAGA CAGAGTTCCA AGGTTGGATT GTACTTGTCG TTTTCAATCA TATTGATAGT	3360
CTCTCTCGAG ACACCGATAT CCTTGGCGAG TCGAGCTGG GAAATACCCA ATTCCTTGCG	3420
AAATCTTTTC ACACGATTCA TCTGTTCTCC TTTCTGATTT ATGTCGTATA TATTTGACTA	3480
TATTATAGTC TTTTAAACAT AAAGTGTCAG GTATTTTTGA CATATTTTTT GAAGAAATAG	3540
TAGTCTCCTT GTCCTATTTG TCTGACAAGT GCAAGCTGGT CGGATTTGTG GTAAAAATAG	3600
TAAGATATGA CAAAAGAATT TCATCATGTA ACGGTCTTAC TCCACGAAAC GATTGATATG	3660
CTTGACGTAA AGCCTGATGG TATCTACGTT GATGCGACTT TGGGCGGAGC AGGACATAGC	3720
GAGTATTTAT TAAGTAAATT AAGTGAAAAA GGCCATCTCT ATGCCTTTGA CCAGGATCAG	3780
AATGCCATTG ACAATGCGCA AAAACGCTTG GCACCTTACA TTGAGAAGGG AATGGTGACC	3840
TTTATCAAGG ACAACTTCCG TCATTTACAG GCATGTTTGC GCGAAGCTGG TGTTCAAGAA	3900
ATTGATGGAA TTTGTTATGA CTGGGAGTG TCTAGTCCTC AATTAGACCA GCGTGAGCGT	3960
GGTTTTTCTT ATAAAAAGGA TGCGCCACTG GACATGCGGA TGAATCAGGA TGCTAGCCTG	4020
ACAGCCTATG AAGTGGTGAA CAATTATGAC TATCATGACT TGGTTCGTAT TTTCTCAAG	4080
TATGGAGAGG ACAAAATCTC TAAACAGATT GCGCGTAAGA TTGAGCAAGC GCGTGAAGTG	4140
AAGCCGATTG AGACAACGAC TGAGTTAGCA GAGATTATCA AGTTGGTCAA ACCTGCCAAG	4200
GAACTCAAGA AGAAGGGGCA TCCTGCTAAG CAGATTTTCC AGGCTATTCT AATTGAAGTC	4260
AATGATGAAC TGGGAGCGGC AGATGAGTCC ATCCAGCAGG CTATGGATAT GTTGGCTCTG	4320
GATGGTAGAA TTTCAGTGAT TACCTTTTAT TCCTTAGAAG ACCGCTTGAC CAAGCAATG	4380
TTCAAGGAAG CTTCAACAGT TGAAGTTCCA AAAGGCTTGC CTTTCATCCC AGATGATCTC	4440
AAGCCCAAGA TGGAATTGGT GTCCCGTAAG CCAATCTTGC CAAGTGCGGA AGAGTTAGAA	4500
GCCAATAACC GCTCGCACTC AGCCAAGTTG CGCGTGGTCA GAAAAATTCA CAAGTAAGAG	4560
GGAAAAAGAT GGCAGAAAAA ATGGAAAAAA CAGGTCAAAT ACTACAGATG CAACTTAAAC	4620
GGTTTTTCGG TGTGGAAAAA GCTTTTTTACT TTTCATTGC TGTAACCACT CTTATTGTAG	4680
CCATTAGTAT TATTTTTATG CAGACCAAGC TCTTGCAAGT GCAGAATGAT TTGACAAAAA	4740
TCAATGCGCA GATAGAGGAA AAGAAGACCG AATTGGACGA TGCCAAGCAA GAGGTCAATG	4800

223

AACTATTACG TGCAGAACGT TTGAAAGAAA TTGCCAATTC ACACGATTTG CAATTAAACA	4860
ATGAAAATAT TAGAATAGCG GAGTAAGATA TGAAGTGGAC AAAAGAGTA ATCCGTTATG	4920
CGACCAAAAA TCGGAAATCG CCGGCTGAAA ACAGACGCAG AGTTGGA AAA AGTCTGAGTT	4980
TATTATCTGT CTTTGTTTTT GCCATTTTTT TAGTCAATTT TCGGTCATT ATTGGGACAG	5040
GCACTCGCTT TGGAACAGAT TTAGCGAAGG AAGCTAAGAA GGTTTCATCA ACCACCCGTA	5100
<hr/>	
CAGTTCCTGC CAAACGTGGG ACTATTTATG ACCGAAATGG AGTCCCGATT GCTGAGGATG	5160
CAACCTCCTA TAATGTCTAT GCGGTCATTG ATGAGAACTA TAAGTCAGCA ACGGGTAAGA	5220
TTCTTTACGT AGAAAAACA CAATTTAACA AGGTTGCAGA GGTCTTTCAT AAGTATCTGG	5280
ACATGGAAGA ATCCTATGTA AGAGAGCAAC TCTCGCAACC TAATCTCAAG CAAGTTTCCT	5340
TTGGAGCAAA GGGAAATGGG ATTACCTATG CCAATATGAT GTCTATCAAA AAAGAATTGG	5400
AAGCTGCAGA GGTCAAGGGG ATTGATTTTA CAACCAGTCC CAATCGTAGT TACCCAAACG	5460
GACAATTTGC TTCTAGTTTT ATCGGTCTAG CTCAGCTCCA TGAAAATGAA GATGGAAGCA	5520
AGAGCTTGCT GGGAACCTCT GGAATGGAGA GTTCCTTGAA CAGTATTCCT GCAGGGACAG	5580
ACGGCATTAT TACCTATGAA AAGGATCGTC TGGGTAATAT TGTACCCGGA ACAGAACAAG	5640
TTTCCCAACG AACGATGGAC GGTAAAGGATG TTTATACAAC CATTTCCAGC CCCCTCCAGT	5700
CCTTTATGGA AACCCAGATG GATGCTTTTC AAGAGAAGGT AAAAGGAAAG TACATGACAG	5760
CGACTTTGGT CAGTGCTAAA ACAGGGGAAA TTCTGGCAAC AACGCAACGA CCGACCTTTG	5820
ATGCAGATAC AAAAGAAGGC ATTACAGAGG ACTTTGTTTG GCGTGATATC CTTTACCAAA	5880
GTAACATGA GCCAGGTTCC ACTATGAAAG TGATGATGTT GGCTGCTGCT ATTGATAATA	5940
ATACCTTTCC AGGAGGAGAA GTCTTTAATA GTAGTGAGTT AAAAAATTGCA GATGCCACGA	6000
TTGAGATTG GGACGTTAAT GAAGGATTGA CTGGTGGCAG AACGATGACT TTTTCTCAAG	6060
GTTTTGCACA CTCAAGTAAC GTTGGGATGA CCTCCTTGA GCAAAAGATG GGAGATGCTA	6120
CCTGGCTTGA TTATCTTAAT CGTTTTAAAT TTGGAGTTCC GACCCGTTTC GGTGTGACGG	6180
ATGAGTATGC TGGTCAGCTT CCTGCGGATA ATATTGTCAA CATTGCGCAA AGCTCATTTG	6240
GACAAGGGAT TTCAGTGACC CAGACGCAAA TGATTCGTGC CTTTACAGCT ATTGCTAATG	6300
ACGGTGTCAT GCTGGAGCCT AAATTTATTA GTGCCATTTA TGATCCAAAT GATCAAACCTG	6360
CTCGGAAATC TCAAAAAGAA ATTGTGGGAA ATCCTGTTTC TAAAGATGCA GCTAGTCTAA	6420
CTCGGACTAA CATGGTTTTG GTAGGGACGG ATCCGGTTTA TGAACCATG TATAACCACA	6480
GCACAGGCAA GCCAACTGTA ACTGTTCTCTG GGCAAAATGT AGCCCTCAAG TCTGGTACGG	6540

224

CTCAGATTGC TGACGAGAAA AATGGTGGTT ATCTAGTCGG GTTAACCGAC TATATTTTCT	6600
CGGCTGTATC GATGAGTCCG GCTGAAAATC CTGATTTTAT CTGTGTATGT ACGGTCCAAC	6660
AACCTGAACA TTATTCAGGT ATTCAGTTGG GAGAATTTGC CAATCCTATC TTGGAGCGGG	6720
CTTCAGCTAT GAAAGACTCT CTCAATCTTC AAACAACAGC TAAGGCTTTA GAGCAAGTAA	6780
GTCAACAAAG TCCTTATCCT ATGCCTAGTG TCAAGGATAT TTCACCTGGT GATTTAGCAG	6840
AAGAATTGCG TCGCAATCTT GTACAACCCA TCGTTGTGGG AACAGGAACG AAGATTAAAA	6900
ACAGTTCTGC TGAAGAAGGG AAGAATCTTG CCCCGAACCA GCAAGTCCTT ATCTTATCTG	6960
ATAAAGCAGA GGAGGTCCA GATATGTATG GTTGGACAAA GGAGACTGCT GAGACCCCTG	7020
CTAAGTGGCT CAATATAGAA CTGAATTTC AAGGTTCTGG CTCTACTGTG CAGAAGCAAG	7080
ATGTTCTGTC TAACACAGCT ATCAAGGACA TTAATAAAAT TACATTAACT TTAGGAGACT	7140
AATATGTTTA TTCCATCAG TGCTGGAATT GTGACATTTT TACTAACTTT AGTAGAAATT	7200
CCGGCCTTTA TCCAATTTTA TAGAAAGGCG CAAATTACAG GCCAGCAGAT GCATGAGGAT	7260
GTCAAACAGC ATCAGGCAAA AGCTGGGACT CCTACAATGG GAGGTTTGGT TTCTCTGATT	7320
ACTTCTGTTT TGGTTGCTTT CTTTTTCGCC CTATTTAGTA GCCAATTCAG CAATAATGTG	7380
GGAATGATTT TGTTTCATCTT GGTCTTGTAT GGCTTGGTCG GATTTTTAGA TGACTTTCTC	7440
AAGGTCCTTC GTAAAATCAA TGAGGGGCTT AATCCTAAGC AAAAATTAGC TCTTCAGCTT	7500
CTAGGTGGAG TTATCTTCTA TCTTTTCTAT GAGCGCGGTG GCGATATCCT GTCTGTCTTT	7560
GGTTATCCAG TTCATTTGGG ATTTTTCTAT ATTTCTCTCG CTCTTTTCTG GCTAGTCGGT	7620
TTTTCAAACG CAGTAAACTT GACAGACGCT GTTGACGGTT TAGCTAGTAT TTCCGTGTGT	7680
ATTAGTTTGT CTGCCTATGG AGTTATTGCC TATGTGCAAG GTCAGATGGA TATTCTTCTA	7740
GTGATTCTTG CCATGATTGG TGGTTTGCTC GGTTCCTTCA TCTTTAACCA TAAGCCTGCC	7800
AAGGTCCTTA TGGGTGATGT GGAAGTTTG GCCCTAGGTG GGATGCTGGC AGCTATCTCT	7860
ATGGCTCTCC ACCAAGAATG GACTCTCTTG ATTATCGGAA TTGTGTATGT TTTTGAAACA	7920
ACTTCTGTTA TGATGCAAGT CAGTTATTTC AACTGACAG GTGGTAAACG TATTTTCCGT	7980
ATGACGCCTG TACATCACCA TTTTGAGCTT GGGGGATTGT CTGGTAAAGG AAATCCTTGG	8040
AGCGAGTGA AGGTTGACTT CTCTTTTGG GGAGTGGGAC TTCTAGCAAG TCTCCTGACC	8100
CTAGCAATTT TATATTTGAT GTAAGAATGG CACCCTGATG TTTCAGGG	8148

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9909 base pairs  
 (B) TYPE: nucleic acid

225

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

TACTCCACCC TTAATATCCG TTCCTGTAAA TACTTTACCG CTTTAAAGTT CATAGAATTG	50
AACTTTTAAA TGCTTGCTT CAAGCATCTT TTCCATCCAA TTTTtaggag TTTGACCAGC	120
TTTAAATAAA AACCTTGCTG GGGTGATTAG TATAGATTTA TCTGCGATTT TATAAGCTTC	180
ATCAATAAAA TAGTGATATA TCGGCTCATC TCTGGCTTCT CCTGTTTCCT GATACGGAGG	240
ATTTCTATC ACGACATCAA ATTTcatttc ACTTTCCTCG CTAGATAGGC GCTCAAAACC	300
TATCATCTA TTCTTTTCC AGTCTTTGAT ATGGGTTTTA GATTCTTCTA CTTCTGGAC	360
TTCTAGCTCA TCCGCAAACA AACTCAATTG TTGAGATTGC TTTTgTTTtag CTGAATAAGG	420
ACTACTTTTT TTCAATCCAT CCATCTGAAA GACATTGTAA GAGATAATAG TCGCAATTTc	480
TTTCTTTTGC TCTAATGTTG GTTGATTTC AGTCTTAGCT AGATAATAGT CCTCAAAAGT	540
TGCCAAAAGA TTCTCACGGC CCAAAGGAG AGAATCTCCT TGATACTCAT AACCATACGA	600
AGCATGATAA GCATCTTTTA CAAGTTTATA AAATGTGACT TCATCTGAAA CCTCAGGACT	660
AATCCGTTGC AGTTTTCTAT CAACAAAACC AACTCGCTCA GATAATGGAA TTTCTCACC	720
AGTTACGGTA TCATATCTCG TTACCATATA AGGTGCTTCA CCACAAGTTA CCTCTAACCA	780
TCGTAAGTCC ACATACTCCT CAAGACTTAA CGAGCCTAAT TTCGATTCTA CATATCCATT	840
TTGCTTTGCG ACCAACCACG TTGGTGTAAG CACTTCTGCC CTTATTTTTG TCCGATCTTT	900
TTGTTTCATAT TTGGATTTTT CAGATCTGGG CTGAATCAAG TTGGCAAAGT TTCCAGTAAC	960
CTTACTTGGA TTGATGCGAT CACTTGGAGC AAATCCCTTT CCTAACAATT CATAAGAATG	1020
CGTAnGCCAA ACAATTGATT TCTTTGTCGT TCGATCTTTT AAAAGAATTT TTAATAAGTC	1080
AGCCGATTCT TTAGCCAAAC TTTCTTCACT AATATCTATT GTCATCAGCA ACCTCTCTTA	1140
TATTGTAAGC CCTATTATAT CATATTTTAA AGAATGAAAA TTTACTTGAA AAAAGTAATT	1200
CAATAAATAT CTCCTCGATG ACCAACTTCT AGAGTAGCAA CGACTAATTC ATCATCTACA	1260
ATTTGTACGA TAACTCGATA ATTACCAATT CTATAGCGCC ATTGACCAAC GCGATTACCA	1320
ACCAAAGCCT TTCCGTGTCG TCTTGGGTCT TCCAAACAT TGGTTTGTA ATAGTTTGTA	1380
ATTAGCTTCT GCGTATAACG GTCCAATTTT TTCAATTGCT TGATAAAACG TCTTGTGGA	1440
ACTAATTTAT ACAAATTATT CATCCTTCAA GCCTAAATCA TGCATCATTT CTTCCCAAGT	1500
AATGGGTTCA ACTCCTTTTT CCAAGTCTTC TAAATACTCT TGATAGGCTA AATCTGCCAC	1560



226

ACCAGCATCG TATTCATCTT CTAGGGCTTC AAGAGTTTGT GTGCGAATAA GTTCCGAAAG	1620
GGAAACTCCT TCAAACCTAG CCATTGCTTT CATAAATGTT TTATCAGCTT CAGAAACTTT	1680
TAATGTAATA GTAGTCATCT TTTGTGCTCC CTTTTTTAAT GGTAACACCA TTGTATTACT	1740
TTTTAGGTGT TCAGTCAATA TAAAAAGAAC ACCTTCTCAG CGTTCTTTCT ATATCTCTGT	1800
CAATGGTGT GCGGTATCTG GTGAGGTATC ATAAACCTTA AAGTCTACTC CGACTCCCAG	1860
ATCAGCTTGA GCCAGCTGAT TGACCATGGT CATATGAGCC AGTTCCTTGA TATTGTTTTC	1920
CTTAGATAAA TGCCCAAGGT AAATCTTCTT AGTACGATTT CCTAGCGTCC GAATCATAGC	1980
TTCAGCACCG TCCTCGTTAG AAAGGTGACC AAGGTGAGAT AGGATTCTGT GTTTGAGTCG	2040
CCAAGCGTAA GAACCTGATC GCAAAATCTC TACATCATGG TTGGCCTCGA TAAGATAACC	2100
ATCCGCATTT TCGACAATGC CCGCCATACG GTCACTGACA TAACCTGTAT CTGTCAAGAG	2160
GACAAAACTC TTATCATCCT TCATAAAGCG ATAGAACTGC GGTGCGACTG CATCATGGCT	2220
TACACCAAAA CTCTCGATGT CGATATCTCC AAAGGTTTTG GTTTTACCCA TTTCAAAAAT	2280
ATGCTTTTGC GAAGAATCCA CCTTGCCAAG ATATTTACTA TTTTCCATAG CTTGCCAGGT	2340
CTTTTCATTG GCATAAAGAT CCATACCATA CTTGCGAGCC AAAACGCCTA CTCCATGGAT	2400
ATGATCTGAA TGCTCATGGG TAATCAAGAT GGCATCCAGG TCTTCTGGCT TACGGTTAAT	2460
TTCAGCTAGC AGACTGGTAA TTTTCTTGCC AGACAAGCCT GCATCTACTA AAAGCTTCTT	2520
TTTTGAGGTT TCCAGATAAA AAGAATTTCC ACTGGAACCC GACGCTAAAA TACTGTATTT	2580
AAAGCCTATT TCACTCATTC TAGTCTTCTA CTTATCCTC CCATACTTCT TCTTTCACCTG	2640
CATCCTTATC ATAAGGGAGT ACAATGGTAA AGGTTGAACC CTTGCCGTAT TCACTCTTGG	2700
CCCAAATAAA GCCCTTATGT TGTTTGATAA TTTCTTTAGC GATAGACAGT CCTAGACCTG	2760
TACCACCTTG TGCACGACTT CTAGCACGAT CCACACGATA GAAACGGTCA AAGATACGTG	2820
GTAAATCCTG CTTAGGAATC CCCAAACCGT GGTGAGAAAT GGATAAAATC ATCTGGTCTT	2880
CAGTTGTCTT CATCTGACA GTGATTTTAC CCCCATCTGG CGAATACTTA ATAGCATTAT	2940
TTAAATATT GTCGACAACC TGCGTCATCT TATCTGTATC AATTTCCATC CAGATAGAAT	3000
TGATGGGATA ATCTCTCACC AACTCATATT TTTTCTCCTT TTCCTGTCCT TTCATCTTGT	3060
CAAAACGATT GAGGATAAAG GTAATAAAG CAGTGAAGTT AATCAGTTCC ACATCTAGGT	3120
GACTGGTAGC ATTATCAATA CGTGAAAGAT GGAGGAGATC CGTCACCATG CGCATCATAC	3180
GGTTGGTCTC ATCAAGAGAA ACCTTGATAA AGTCTGGTGC TACAGTTTCA CACAAAGCCC	3240
CCTCATCCAA GGCTTCAAGA TAGGATTTTA CGCTAGTCAG AGGAGTCCGT AACTCATGGC	3300
TAACATTGGA AACAAAGAGT CTTGCTTCGC GTTCTTCCTT CTCCTGCTCC GTCGTATCAT	3360

227

GCAAAACAGC CACCAAACCT GAAATAAAGC CAGACTCTCG ACGTATCAAG GCAAAGCGAA	3420
CTCGAAGGTT CAAATATTCG CCATTGATAT CTTGGGAATC TAGCAACAAT TCTGGACTTT	3480
GGTAATCAA ATCACGCAAT TCATAGTTTT CTTCTATCTT GAGCAATTC AAAATGCTTC	3540
TATTCAGAAC ATCTTCCTTA ACCAACCCCA GTTGCTTCTT GGCTGTATCG TTAATCATGA	3600
TAATCTGACC CCGACGGTTA GTCGCAAGAA CCCCATCTGT CATATAAAAC AGAATACTAT	3660
TTAGCCTCTT ACTCTCTGT TCTAGATTTT CCTGAGTGAG ACGAATAACC TCCGACAAGT	3720
CATTCAAATT ATTGGTAATA TTGGTGATTT CAGACCCACC TTGCATATCA AGAACCTTGG	3780
AATAATCTCC TGCAATCAAA TCTTTAACCT TTTGATTGAC TTGCTTCAAC TGAATATTAT	3840
CACGTCTATT TTCCAGTAAT AAGAGGGTCA CAACAAGGAT GAAACCTAAC AAAATCAGGA	3900
TAAAGATAAA ATCTCTGGTA AAAATGGTTT GTTTCAGTAA ATCAAGCATT ATTTCTCATG	3960
TAATACCCTA CACCACGGCG CGTCAAGATA TACTCTGGTC GGCTGGGCGT ATCTTCAATC	4020
TTCTCAGCA GACGTCGTAC AGTCACATCA ACTGTACGGA CATCACCAA ATAGTCATAA	4080
CCCCAGACAG TCTCAAGCAA GTGTTGCGCG GTGATGACTT GACCTGTATG CGATGCTAAA	4140
TGATACAAAA GCTCAAATTC ACGATGGGTT AAGTCTAGTT CTTGCCATA TTTTITAGCC	4200
ACGTAGGCGT CTGGAACAAT TTCTAAATCC CCAATTTGGA TAGGTTGAGG TTTACTATCT	4260
GCTTCCTGAC CATCTACTGG CATAGGTTGA GAACGACGCA GAAGAGCTTT AACACGCGCC	4320
TGCAACTCAC GATTGGAGAA GGGTTTTGTT ACATAGTCAT CTGCCCCAAG TTCCAAACCG	4380
ATAACCTTAT CAAATTCACT ATCTTTGGCT GAAAGCATAA GAATGGGCAC ACTGCTTGTC	4440
TTACGAATGG TCTTAGCAAC TTCTAAACCA TCAATTTCTG GAAGCATCAA ATCCAGAATA	4500
ATAATATCTG GTTGCTCTGC TTCAAATGTC TCTAGCGCTT CACGACCATT AAAAGCAGTT	4560
ACAACCTTCGT AACCTTCCTT GGTCATATTA AACTTGATAA TATCCGAGAT TGGTTTCTCA	4620
TCATCTACAA TTAGTATTTT TTTCATATGT TCACCTTTTT CTCTACTATT ATACCAAAAA	4680
AATAGTCAGA AGACACAATA GCTAGTCTTG GCTACTGTCT AAGTTGGCTT GTGCATAAAC	4740
CTGCCAGATT TTTTGTGGG GTTTGGCAAG TGGGTAATTC TTGAATTCTT CTGGTGAAAG	4800
CCAGCGAACT TCCCTATCTG AAAAATCATG GAAGTCACTC ACCTGACCTG CTACAATCTG	4860
TACATGCCAT TTTGATGAC TAAAAACATG CTGGACTGTA TCAAAACAAA CATCAAGCCA	4920
ATCAACATCT AGGTCATAGT CCTGCTGGAA ACTCTCTTCT GGACTGGGAC CAAAGTTTAC	4980
ACTTTCTTCC GCAACCTGAT GAAAGAGGTC AACTGCTCT TCTTGCGAAA AGTTATCAAC	5040
TTCTATAAAG GGGAAATGCC AAAAACCTGC CAAGAGCTTT TCGCTTTCAT TTTTTTCAAG	5100

228

TAAAAATTGT CCTTGAGAAT TTTTCACAAC TAAGGCTTTA AGATAAATAG GAACCGGCTT	5160
TTTCTTAGGA GATTTAATTG GATAACGGTC CATGGTTCCA TTCTGATATG CCGCACTAAA	5220
GTCCTTGACT GGGCTTTCTT CAGGTCTGGG ATTTACAGGA GACTCAATAT CAGACCCTAA	5280
GTCCATCAAG GCTTGATTAA AATCACC CGG ACGATCCGGA TTAATCAAGA TCTCCATCAT	5340
TGCCTGAAAA ATTTTTCGAT TACTTGGAAT CCCAATATCG TGGTTGACTT CAAACAGACG	5400
CGCCAAGACC CGCATGACAT TACCATCTAC AGCTGGCTCA GGCAAGTTAA AAGCAATACT	5460
GGAAATGGCT CCTGCTGTGT AAGGTCCAAT CCCTTTCAAG CTGGAAATTC CTTCATAGGT	5520
ATTTGGAAAT TGGCCACCAA AGTCAGTCAT AATCTGCTGG GCTGCAGCCT GCATATTGCG	5580
AACTCGAGAA TAATAGCCCA AGCCCTCCCA AGCTTTCAGT AAACCTCTCT CAGGCGCAGT	5640
TGCCAGACTT TCGACAGTTG GAAACCAGTC CAAAAATCTT TCGTAGTAAG GGATAACTGT	5700
ATCCACCCTG GTCTGCTGAA GCATGATTTC AGATACCCAG ATGTGATAAG GATTTTACT	5760
TCTCTCCAA GGCAAACTC TTTTGTTC ATCATACCAA GCGAGAAGTT TCTCACGGAA	5820
AGAAATGACT TTCTCTCCG GCCACATGAC GATACCGTAT TCTTCAAAT CTAACATATC	5880
TCTAGTATAA CACAGAAGGT TTCACCTGTC TTTGTATCTG ATTTATAATA TTTCAATAG	5940
ATAGTATATA ACTTTTCTAT CTACTTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG	6000
CTAGCCGCAG GTTGCTCAAA ACAGTGTTC GAGGTGTGG ATAGAACTGA CAGAGTCAGT	6060
ATCATATACT ACGGCAAGGT GAAGCTGACG TAGTTTGAAG AGATTTTCGA AGAGTATAAA	6120
TCTTATTGAT GAACTGCTTG CAGTCTGAGA AAAAATGAGC TTGGATATTA TTTCCAACT	6180
CACCTAAAGT CAATTTCAAT CCACTAGAAC AAGCCTAGTA CAGTTCATC GCTTCAACA	6240
TCCA7GTTGA GAGCTGCTGG ACGTTTGGGA AGACCTGGCA TGGTCATAAC ATCACCAGTT	6300
AAGGCAACGA TGAAGCCTGC ACCTAATTTT GGTACCAATT CACGAATGGT AATTTCAAAG	6360
TTTCTGGTG CTCCAAGCGC ATTTGGATTG TCTGAGAAAC TGTATTGAGT TTTAGCCATA	6420
CAGATTGGCA ATTTGTCCCA ACCGTTTGA ACGATTTGAG CAATTTGTGT TTGAGCTTTC	6480
TTCTCAAAGT TCACTTTGCT ACCACGATAG ATTTCACTGA CAATTTTTC AATCTTTTCT	6540
TGGACAGAAA GGTCAATTATC ATACAAACGT TTATAGTTAG CTGGATTTTC AGCAATTGTC	6600
TTAACAACTG TTTGGCAAG TGCTACTCCA CCTTCTGCTC CATCAGCCCA GACACTAGCC	6660
AATTCAACTG GTACATCGAT TGAGGCACAG AGTTCTTTTA AGGCTGCAAT TTCAGCTTCT	6720
GTATCAGATA CAAATTCGTT AATAGCTACA ACTGCTGGAA TACCGAATT ACGGATATTT	6780
TCAACGTGGC GTTCAAGTT AGCAAAACCT GCACGAAGT CCTCTACATT TTCTTCAGTC	6840
AGAGCGTCTT TAGCCACACC ACCATTATC TTAAGGGCAC GAAGGGTTC GACAATAACA	6900

229

ACTGCATCTG GAGATGTTGG CAAGTTTGGT GTCTTGATAT CAAGGAATTT CTCAGCACCA	5960
AGGTCCGCAC CAAAACCAGC TTCAGTAACA GTGTAATCAG CCAAGTGAAG GGCTGTTGTC	7020
GTCCGCAAAA CAGAGTTACA GCCATGAGCG ATATTGGCAA ATGGACCACC GTGTACAAAG	7080
GCAGGTGTAC CGTAAATTGT CTGAACCAAG TTTGGCTTAA TAGCATCCTT CAAAATCAAA	7140
GCCAAGGCAC CCTCAACCTG CAAATCACCT ACAGAAACAG GCGTACGGTC ATAGCGATAA	7200
CCAATAACGA TATTGCGCAA ACGACGTTTC AAGTCCTCGA TGTCCGTTGC CAAGCAAAGA	7260
ATTGCCATGA TTTCTGAAGC AACTGTAATA TCAAAACCAT CCTCACGTGG AATACCGTTT	7320
AGAGGACCAC CAAGACCAAC AGTCACATGG CGGAGCGTAC GGTGCTTCAA GTCCACAACG	7380
CGTTTCCAGA GGATACGACG TTGATCAATT CCCAGCTCAT TCCCTTGGTG CAAGTGGTTG	7440
TCAATCAAGG CAGAAAGGGC ATTGTTGGCA GTTGTAATAG CATGCATATC TCCAGTAAAG	7500
TGGAGGTTGA TGTCTTCCAT TGGCAGAACT TGTGCATACC CACCACCAGC AGCACCACCC	7560
TTGATCCCCA TGA CTGGACC AAGAGACGGT TCGCGGATAG CAATCATGGT TTTCTTGCCA	7620
ATCTTGTTCA AGGCATCCGC AAGACCAATG GTAAGCGTCG ACTTTCCTTC ACCTGCAGGT	7680
GTTGGGTTGA TGGCAGTAAC CAAGATCAAT TTACCGACTG GATTGCTCTC AACTGCACGA	7740
ATTTTATCAA AGCTGAGTTT AGCCTTGTAAC TTTCCGTACA ACTCCAAATC GTCATAAGAA	7800
ATACCAAGTT TCTCTACAAC ATCAACAATT GGCTTCAACT CAATACTCTG TCGGATTTCA	7860
ATATCTGTTT TCATTCAAAA TTCCTCTAAC CTCTTATATG ATAATTCATT ATATCACAAA	7920
ACAAGATTTT TAACATCCTA AAACCTCTCTA AACGTTCTGA AATATCTCTG TTTTAAAGAC	7980
TTTATAGATC CTTTCTTAAA TTTTATATGG CTTTATAGTT TGAAACTATA ATAAATCTTC	8040
GTTTTTACCA AAAATTTATC ACTTTCATTT TACTTACCGC TTATTTTTGT GTACAATAGT	8100
GCTATGAAAA TTTTAGTTAC ATCGGGCGGT ACCAGTGAAG CTATCGATAG CGTCCGCTCT	8160
ATCACTAACC ATTCTACAGG TCACTTGGGG AAAATTATCA CAGAGACTTT GCTTCTTGCA	8220
GGGTATGAAG TTTGTTTAAT TACGACAAAA CGAGCTCTGA AGCCAGAGCC TCATCCTAAC	8280
CTAAGTATTC GAGAAATTAC CAATACCAAG GACCTTCTAA TAGAAATGCA AGAACGTGTT	8340
CAGGATTATC ACGTCTTGAT CCACTCAATG GCTGTTTCTG ACTACACTCC TGTTTATATG	8400
ACAGGGCTTG AGGAAGTTCA GGCTAGCTCC AATCTAAAAG AATTTTAAAG CAAGCAAAAT	8460
CATCAGGCCA AGATTTCTTC AACTGATGAG GTTCAGGTTT TGTTCCCTAA AAAGACACCC	8520
AAAAATCATAT CCTAGTCAA GGAATGGAAT CCTACTATTC ATCTGATTGG ITTCAAACCTG	8580
CTGGTTGATG TTACCGAAGA TCATCTGGTT GACATTGCAC GAAAAAGTCT TATCAAGAAT	8640

230

CAAGCAGATT TAATCATCGC GAATGACCTG ACTCAAATTT CAGCAGATCA GCACCGAGCT	8700
ATATTGTTG AGAAAAATCA GCTTCAAACA GTCCAGACTA AAGAAGAAAT TGCAGAACTC	8760
CTCCTTGAAA AAATTCAAGC CTATCATTCT TAGAAAGGAA AACTATGGCA AACATTCTCT	8820
TGGCTGTAAC GGGTTCAATC GCCTCTTATA AGTCGGCAGA TTTAGTCAGT TCTCTAAAAA	8880
AACAAGGCCA TCAAGTCACT GTCTTAATGA CTCAGGCTGC TACAGAGTTT ATCCAACCTT	8940
TGACACTACA GGTACTCTCA CAGAATCCTG TCCACTTGGA TGTCATGAAG GAACCCATATC	9000
CTGATCAGGT CAATCATATC GAACTTGGA AAAAAGCAGA TTTATTTATC GTGGTACCTG	9060
CAACTGCTAA CACTATTGCA AAAGTAGCTC ACGGATTTGC GGACAACATG GTAACCAGTA	9120
CAGCTCTAGC CCTACCAAGT CATATTCCCA AACTAATAGC TCCTGCTATG AATACAAAAA	9180
TGTATGACCA TCCAGTAACT CAGAATAATC TGAAAACATT AGAAACTACG GCTATCAGCT	9240
GATTGCTCCT AAGGAATCCC TACTAGCTTG TGGAGACCAC GGACGAGGAG CTTTAGCTGA	9300
CCTCACAATT ATTTTAGAAA GAATAAAGGA AACTATCGAT GAAAAAACGC TCTAATATTG	9360
CACCCATTGC TATCTTTTTT GCTACCATGC TCGTGATACA CTTTCTGAGC TCACCTATCT	9420
TTAACCTTTT TCCATTTCCA ATCAAACCGA CCATTGTTCA TATTCCTGTC ATTATTGCCA	9480
GCATTATTTA TGGTCCACGA GTTGGGGTTA CACTTGGATT TTTGATGGGA TTAAGTAGCT	9540
TGACGGTTAA CACGATTACG ATTCTACCGA CAAGCTACCT CTTCTCTCCC TTCGTACCAA	9600
ACGGAAACAT CTAATCAGCT ATCATTGCCA TCGTCCCACG TATTTTGATT GGTTTAACTC	9660
CTTACTTAGT CTATAAAGT ATGAAAAACA AGACTGGTCT GATTTTAGCT GGAGCCCTTG	9720
GTTCCTTGAC AAATACTATC TTTGTCCTTG GAGGAATCTT CTTCTATTTT GGAAATGTTT	9780
ATAATGAAA TATCCAACCT CTTCTGGCAA CCGTTATCTC AACAAATCCA ATTGCTGAAT	9840
TGGTCATTTT TGCAATTCTA ACCCTAGCCA TTGTTCCACG ACTACAAACC TTGAAAAAAT	9900
AAAAACAGG	9909

## (2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1126 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TAATTTTCAT ATAATAGTAA AATAGAATGT GTGATTCAAT AATCACCTCA AATAGAAAGG	60
AAATTCTATG TCAAACTCTAT CTGTTAATGC AATTCGTTTT CTAGSTATTG ACGCCATTAA	120

231

TAAAGCCAAC TCAGGTCATC CAGGTGTGGT TATGGGAGCG GCTCCGATGG CTTACAGCCT	180
CTTTACAAAA CAACTTCATA TCAATCCAGC TCAACCAAAC TGGATTAACC GCGACCGCTT	240
TATTCTTTCA GCAGGTCATG GTTCAATGCT CCTTTATGCT CTTCTTCACC TTTCTGGTTT	300
TGAAGATGTC AGCATGGATG AGATTAAGAG TTTCCGTCAA TGGGGTTCAA AAACACCAGG	360
TCACCCAGAA TTTGGTCATA CCGCAGGGAT TGATGCTACG ACAGGTCCTC TAGGGCAAGG	420
GATTTCAACT GCTACTGGTT TTGCCAAGC AGAACGTTTC TTGGCAGCCA AATATAACCG	480
TGAAGGTTAC AATATCTTTG ACCACTATAC TTACGTTATC TGTGGAGACG GAGACTTGAT	540
GGAAGGTGTC TCAAGCGAGG CAGCTTCATA CGCAGGCTTG CAAAACTTG ATAAGTTGGT	600
TGTTCTTTAT GATTCAAATG ATATCAACTT GGATGGTGAG ACAAAGGATT CCTTTACAGA	660
AAGTGTTGCT GACCGTTACA ATGCCTACGG TTGCATACT GCCTTGTTG AAAATGGAAC	720
AGACTTGGA GCCATCCATG CTGCTATCGA AACAGCAAAA GCTTCAGGCA AGCCATCTTT	780
GATTGAAGTG AAGACGGTTA TTGGATACGG TTCTCCAAAC AAACAAGGAA CTAATGCTGT	840
ACACGGCGCC CCTCTGGAG CAGATGAAAC TGCATCAACT CGTCAAGCCC TCGGTTGGGA	900
CTACGAACCA TTTGAAATTC CAGAACAAGT ATATGCTGAT TTCAAAGAAC ATGTTGCAGA	960
CCGTGGCGCA TCAGCTTATC AAGCTTGGAC TAAATTAGTT GCAGATTATA AAGAAGCTCA	1020
TCCAGAACTG GCTGCAGAAG TAGAAGCCAT CATCGACGGA CGTGATCCAG TCGAAGTGAC	1080
TCACGACGAC TTCCAGCTT TAGAAAATGG TTTTCTCAA GCAACT	1126

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2520 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCGGCAACAA AAAAGAAAA ATCAACAGTT AAAAAAATC TAGTCATCGT GGAGTCGCCT	60
GCTAAGCCAA GACGATTGAA AAATATCTAG GCAGAACTA CAAGGTTTTA GCCAGTGTCG	120
GGCATATCCG TGATTTGAAG AAATCCAGTA TGTCCGTCGA TATTGAAAT AATTATGAAC	180
CGCAATATAT TAATATCCGA GGAAAAGGCC CTCTTATCAA TGAAGTGAAG AAAGAAGCTA	240
AAAAAGCTAA TAAAGTTTTT CTCGCGAGTG ACCCGGACCG TGAAGGAGAA GCGATTTCTT	300
GGCATTGGC CCATATTCTC AACTTGGATG AAAATGATGC CAACCGTGTG GTCTTCAATG	360

232

AAATCACCAA	GGATGCAGTC	AAAAATGCTT	TTAAAGAACC	TCGTAAGATC	GATATGGACT	420
TGGTCGATGC	CCAACAAGCT	CGTCGGATCT	TGGATCGCTT	GGTAGGGTAT	TCGATTTTCGC	480
CTATTTTGTG	GAAGAAGGTC	AAGAAGGGCT	TGTCAGCAGG	TCGCGTTCAG	TCCATTGCCC	540
TTAAACTCAT	CATTGACCGT	GAAAATGAAA	TCAATGCCTT	CCAGCCAGAA	GAATACTGGA	600
CAGTTGATGC	TGTCTTTAAA	AAGGGAACCA	AACAATTTCA	TGCTTCCTTC	TATGGAGTAG	660
ATGGTAAAAA	GATGAAACTG	ACCAGCAATA	ACGAAGTCAA	GGAAGTCTTG	TCTCGTCTGA	720
CGAGTAAAGA	CTTTTCAGTA	GATCAGGTGG	ATAAGAAAGA	GCGCAAGCGC	AATGCTCCTT	780
TACCTTATAC	CACTTCATCT	ATGCAGATGG	ATGCTGCCAA	TAAAATCAAT	TTCCGTACTC	840
GAAAAACCAT	GATGGTTGCC	CAACAGCTCT	ATGAAGGAAT	TAAATATCGGT	TCTGGTGTTT	900
AAGGTTTGAT	TACCTATATG	CGTACCGATT	CGACTCGTAT	CAGTCCTGTA	GCGCAAAATG	960
AGGCGGCAAG	CTTCATTACG	GATCGTTTTG	GTAGCAAGTA	TTCTAAGCAC	GGTAGCAAGG	1020
TCAAAAACGC	ATCAGGTGCT	CAGGATGCCC	ATGAGGCTAT	TCGTCCGTCA	AGTGTCTTTA	1080
ATACACCAGA	AAGCATCGCT	AAGTATCTGG	ACAAGGATCA	GCTTAAGCTA	TATACCTTA	1140
TCTGGAATCG	TTTTGTGGCT	AGCCAGATGA	CAGCGGCCGT	TTTTGATACC	ATGGCTGTTA	1200
AATTGTCTCA	AAAAGGGGTT	CAATTTGCTG	CCAATGGTAG	TCAGGTAAAG	TTTGATGGTT	1260
ATCTTGCCAT	TTATAATGAT	TCTGACAAGA	ATAAGATGTT	ACCGGACATG	GTTGTTGGAG	1320
ATGTGGTCAA	ACAGGTCAAT	AGCAAACCAG	AGCAACATTT	CACCCAACCG	CCTGCCCCGT	1380
ATTCTGAAGC	AACACTGATT	AAAACCTTAG	AGGAAAAATGG	GGTTGGACGT	CCATCAACCT	1440
ACGCGCCAAC	CATTGAAACC	ATTCAGAAAC	GTTATTATGT	TCGCCTGGCA	GCCAAACGTT	1500
TTGAACCGAC	AGAGTTGGGA	GAAATTGTCA	ATAAGCTCAT	CGTTGAATAT	TTCCAGATA	1560
TCGTAAACGT	GACCTTCACA	GCTGAAATGG	AAGGTAAACT	GGATGATGTC	GAAGTTGGAA	1620
AAGAGCAGTG	GCGACGGGTC	ATTGATGCCT	TTTACAAACC	ATTCTCTAAA	GAAGTTGCCA	1680
AGGCTGAAGA	AGAAATGGAA	AAAATCCAGA	TTAAGGATGA	ACCAGCTGGA	TTTGACTGTG	1740
AAGTGTGTGG	CAGTCCAATG	GTCATTAAAC	TTGGTCGTTT	TGGTAAATTC	TACGCTTGTA	1800
GCAATTTCCC	AGATTGCCGT	CATACCCAAG	CAATCGTGAA	AGAGATTGGT	GTTGAGTGTC	1860
CAAGCTGTCA	TCAGGGACAA	ATTATTGAGC	GAAAAACCAA	GCGTAATCGC	CTATTCTATG	1920
GTTGCAATCG	CTATCCAGAA	TGTGAATTTA	CCTCTTGGGA	CAAGCCTGTT	GGTCGTGACT	1980
GTCCAAAAATG	TGGCAACTTC	CTCATGGAGA	AAAAAGTCCG	TGGTGGTGGC	AAGCAGGTTG	2040
TTTGTAGCAA	AGGCGACTAC	GAGGAAGAAA	AGATGGCTCT	TTGTCAACTG	TAGTGGGTTG	2100
AAGTCAGCTA	AGCTCGAGAA	AGGACAAATT	TTGTCTTTTC	TTTTTTGATA	TTCAGAGCGA	2160

233

TAAAAATCCG TTTTTGAAG TTTTCAAAGT TCCGAAAACC AAAGGCATTG CGCTTGATAA 2220  
 GTTTGATGAG ATTATTGCTC GCTTCCAATT TGGCGTTAGA ATAGTGTAGT TGAAGGGCGT 2280  
 TGACGATTTT CTCTTTGTC TTTAGAAAAG TTTTAAAGAC AGTCTGAAAA AGAGGATGAA 2340  
 CCTGCTTTAG ATTGTCCTCA ATGAGTCCGA AAAATTTCTC CGGTTCTTA TTCTGAAAGT 2400  
 GAAACAGCAA GAGTTGATAG AGCTGATAGT GATGTTTCAA GTCTGTGAA TAGCTCAAAA 2460  
 GCTTGTTTAA AATCTCTTTA TTGGTTAAAT GCATACGAAA AGTAGGGCGA TAAAAATGTT 2520

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10993 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TTTTCTCGAT AATAACTTCC ACCTTATTAT TTGGGATACC CTCCTCTTCT TCACCACCAC 60  
 GTTCATAGTA GTCATCGCGA TAGAGAAAAG CTACGATATC AGCGTCCTGC TCAATAGACC 120  
 CAGATTACAG AATATCAGAC AAGACCGGTC TCTTGTCTTG ACGTTGTTCT ACACCACGAG 180  
 AAAGCTGACT CAGAGCGATT ACTGGAACCT TCAATTCCTT GGCTAGTATT TTCAACTGAC 240  
 GAGAAATTTC AGAAACTTCT TGTGACGAT TTTCTCGACC AGTCCCGTG ATAAGTTGCA 300  
 AATAGTCTAT CAAAATCAAA CCAAGATTTC CAGTTTCTTG AGCCAATTTA CGAGAACGAG 360  
 AACGAATCTC TGTAATCCGA ATACCTGGCG TATCATCGAT ATAGATACTG GCGTTAGCTA 420  
 GATTACCCTG AGCAATAGTA TATTTTGGCC ACTCCTCATC TGTCAATTGC CCTGTACGGA 480  
 TAGAATGTGA CTCCACTAAG CTTTCTGCAG CTAACATACG ATCTACCAAG CTTTCCGCAC 540  
 CCATTTTCGAG TGAAAAAATA GCAACCGTTT TGTCCAACCT AGTCCCAATG TTCTGAGCGA 600  
 TATTCAAGGC AAATGCTGTC TTACCAACTG CTGGACGAGC TGCTAAGATA ATCAACTCCT 660  
 CCTCATGAAG TCCTGTTGTC ATATGATCCA AATCAGGATA ACCTGTCGCA ATACCTGTAA 720  
 TATCGGTCCT TTGTTGCGAG CGAGCTTCCA GATTTCCAAA GTTGAGATTC AACACATCTC 780  
 GAATGTTCTT AAACCCGCTT CGATTTGCAT TTTCACTGAC ATCAATCAAC CTTTTTCTG 840  
 CCTGAGCAAT AATTTATCA GCTGGTTGTG ACGCTTCGTA AGCTTGTTG ACAGACTCTG 900  
 TCAACTTGGC AATTAAACGA CGTAGCATG CTTTTCTGC AACAACTTA GCATAATACT 960  
 CCGCATTAGC AGAAGTTGGC ACAGAATTAA CAATCTCAAC CAAGTAAGAC AAGCCACCAA 1020



234

TATTCGTGTA	ATCACCTTGA	TTATCAAGGA	TAGTACGAAC	CGTTGTTGCA	TCTATGGCAT	1080
CACCACGATC	GGATAAATCG	ACCATGGCTT	GGAAAATCAA	ACGATGGGCA	TACTTAAAAA	1140
AGTCCCGAGA	CTCAATGTAT	TCTCGCACAA	AAACAAGTTT	ACTCTCATCA	ATAAAGATAG	1200
CCCCTAAAC	GGATTGCTCA	GCTAAGATAT	CTTGAGGTTG	TACTCGTAAC	TCTTCTACTT	1260
CTGCCATCAG	ACTTCCCTTC	CTTTTACAAT	CTTGTCAGA	AGGTGTAAAC	TTATCCTTCT	1320
TTACACAGAA	GATTGATTAC	ACTTGTGATA	TCTTGATAGA	TTTTCACTGG	CACATCAATC	1380
AAACCAACCG	CTCGAATCGG	AGCTTGTACT	TGAATATGAC	GTTTATCAAT	CTTAATTCCA	1440
AATTGCTTTT	GCAATCTTTC	TGCAATCTTC	TTATTGGTAA	TAGAACCAAA	GGTACGACCA	1500
TCTGGACCAA	CTTTTCAAC	AAATTCTACA	ACAGTTTCTT	CTGCTTCAAG	TTGTGCTTTA	1560
ATTGCTTTTC	CTTCTGCAAT	CATCTCAGCG	TGAGCTTTTT	CTTCCGATTT	TTGTTTACCA	1620
CGAAGTTCAC	CTACAGCTTG	AGCAGTCGCT	TCTTTGGCTA	GATTCTTTTT	GATAAGAAAG	1680
TTTTGCGCAT	ACCTGTGG	TACTTCCTTA	ATTCGCTT	TTTTACCTTT	TCCTTTAACA	1740
TCTGCTAAAA	AGATTACTTT	CATTCTTCTT	TCTCCTTTTC	CTTCATTTCA	TTTAATACAA	1800
TTTCTGTGAG	TTTTTCACCT	GCTTCTGACA	AGGTTACATC	TTTAATTTGA	GCTGCTGCCA	1860
AATTAAAGTG	GCCTCCACCG	CCTAACTCTT	CCATAATCCG	TTGTACATTC	AGTTTACTAC	1920
GACTTCGAGC	TGAGATAGAG	ATAAATCCTT	GTGTATTCTT	CGCAAGAACA	AAACTCGCTT	1980
CAATACCTGA	CATGGCTAAC	ATGGCATCTG	CTGCCCTACT	AATAACAAC	GTATCATAGC	2040
ATTTATGTC	CTTAGCCTCT	GCTATTAGTA	CATCTGAACC	TAATTTACGC	CCCTGTAAAA	2100
TAAGTTCATT	GACCTCACGA	TATTCCTCAA	AATCTGTCCG	AGCGATTTC	TGGATAGCAA	2160
TACTATCACT	TCCGCGCGTT	CTGAGATAGC	TAGCAACATC	AAATGTCCGA	CTAGTTACTC	2220
GCGAGGTGAA	ATTTTGTAGTA	TCCAACATCA	TACCAGCCAT	CAAGACACTT	GCTTGATAC	2280
GACTCAAACG	ATTTTCTTA	GAATTCTGGA	ACTGAATCAA	TTCCGTTACC	AACTCACTGG	2340
CACTACTTGC	ACCACTTTTCG	ATATAAGTAA	TAACCGCATT	ATCTGGAAAA	TCCTGATCCC	2400
TTCTATGGTG	GTCAATAACA	ATGGTTTGGG	TAAATAAATC	ATAAAATTCT	TTTGATAATG	2460
TTAAGGCTGT	CTTTGAATGG	TCTACAAGAA	TCAACAAAGA	ACGATTGGTC	ACCATCCCCA	2520
TTGCATCCTT	AACAGACAAC	AACTTCGTAA	CTCCTTCTTT	TTCTATGAAT	GAAACAGCTC	2580
GTTCATATC	TGGAGACATT	TGTTCTTCAT	CATAAAGAGC	ATAGCTATTT	TCAATCACAT	2640
TGCTGGCGAA	CAACTGCATA	CCTACAGCAG	AGCCCCAAGC	ATCCATGTCT	AAATTTTGT	2700
GACCGACTAC	AAAAACCTGA	TCTACACTCC	GAATCTTATC	TGAAATAGCT	GTATCATAG	2760
CGCGGTACG	AGTCCGTGTA	CGCTTGATTG	AAGCAGCAGA	CCCACCACCA	AAATAAACTG	2820

235

GATTTTTCGT TTCGTCGTTT TCCTTAACAA CCACCTGGTC GCCACCACGT ACTTCAGCCA	2880
AGTTCAAATT GAGCAAAGCA ACTTTCCTTA TCTCATCATG ATTTCCATCG CCATAAGAAA	2940
ATCCCATACT TAAGGTCAAG GGCAACTGTC TCTGTTTCGA CTCTTCTCTG AAAGCATCAA	3000
TAACAGAAAA TTTATCATTC ATCAAGCCCT CAAGCACCGT GTAGTCAGTA AATAGATAAA	3060
ATCGATCCAT ACTTACCCGA CGAGAAAACA TCATGTGTTT TTCTGAAAAC TCTGATATAA	3120
AATTAGCTAC AAAACTATTG ATTTGACTAA TATCTGACTC AGAAGTTTCA TCCTCCAAAT	3180
CATCATAATT ATCCACAGAG ACAATCCCAA TCACTGGTCT ACTTGTACC AATTCATCTG	3240
TTATGGCTTG TTCCCTGGAT ACATCTACAA AATACAAAC ACCGGAAGAA GCATCCATAT	3300
GAACAGCATA ACGCTTCTCA CCAAGCTTGG CATAAGTAGA CGGATTTCTT ACTGAAGCCT	3360
TGATAATCGT TTGAACAGCT TCTAAATCAA AATCACCATC TTCCTTGGTC AAAATCAATT	3420
CAGCATAGGG ATTAAACCAC TCAACCTCTC CAGAAGATAA ATTCAATTTC ATAACACCTA	3480
CAGGCATCTG TTCCAATAGA GCTGTCAAAC TTTCTCCGC TTGGTGGTTT ACATACTGTA	3540
TCTGTCTAC ATCACTCCTT GTATAATGCA CTCTCAGTTT CTTAAATAAA AAAACATAGC	3600
CTCCTACAAA AAGAAACAAA ATTAAAACCG TCAACAGATT ATTATTAACA AAAATAATGA	3660
AAGTGGATAA GACTCCAAAC GCAATCAATC CTACTAGAAT AGGAAAAAAT GGACTTACAT	3720
AAAATTTTTT CATTCAAAAC CTCTTGGCAC CCATTATACC ATAATACCCC TCAAAAAGCG	3780
ACTTTTTAAA AGTGTAATCA GTAATTCTAT CAATTATAAG AAAAAGGTAG TTTACAATTC	3840
AGTAAACCTA CCTTTACACA TATTGAAATT AAGATTCTTT AACCTCTAAC AAACCAATTT	3900
CGCCATCCTC ACGACGATAA ATCACATTGG TTGTCTGATC TTCAACATCC ACATAGATAA	3960
AGAAATCATG CCCCATAAAA TCCATTTGTA GAATTGCTTC TTCCAAATCC ATTGGTTTTA	4020
AATCAATTTG TTTTGAACGA ACAACTTTAG ACTGGACAAT ATTTGAATCT TCCACCAAAG	4080
CATCTGTAAG TAATTGACCA GTTGCTACCT TATTTTATT TTTACGCTCG ATTTTGTATT	4140
TATTTTACG AATCTGACGT TCAATTTTAT CAGTTACAAG GTCAATTGAA CCATACATAT	4200
CTTGAGATAC ATCTTCTGCG CGGAGAGTAA TAGATCCAAG CGGAATCGTT ACTTCCACTT	4260
TAGCCGTTTT TTCACGATAA ACTTTTAAGT TAATTCGGGC ATCCAACTCT TGTCTGGTT	4320
GGAAGTACTT TTCGATCTTT TCGAGTTTAG AACTACATA ATCACGAATT GCTTCTGTTA	4380
CTTCTAGGTT TTCACCACGG ATACTATATT TAATCATATG AGTACCTTCT TTCTAAACAT	4440
TTTTGTTTTT ATGATTTTAT TATAACGCTT TCATTCTATT TTTGCAAATT TTTTCTCAT	4500
CTTACAAGGG AAAATGTTTT TACATCCTTA GCACCAGCTT CTTCCAACAG TTTCTTAACA	4560

236

CGATTTATAG TTGCTCCTGT AGTATAGATA TCATCTATAA GTAGGATTTT TTTAGGAATA	4620
GTGACTCCAC TTTTAATAAA GAAAGGAAGT TCTGTCCCA AGCGCTCTGA ACGATTTTTA	4680
GAAGAACTGG CTCTCTCTC TCTTTTCTCT AATAAATCCA GATACTCAAA GCCTGCTGCC	4740
TCTACCAAGC CCTCAACCTG ATTAAATCCT CTATTAGCAT ATCTATCAGG ACTTAGGGGA	4800
ATTACAACAA ATTGATACTC TTTGTACTTT TTCAACTCCT CACTTAAAAA TGAAGCGAAA	4860
ACTTTTCTTA ACAGGAAGTC TCCATCAAC TTATACCGAC TGAAAAATC CTTCATAGCT	4920
TGATTGTAAG TAAAAATCGC TCTATGACTG ACTTCAACTC CCTCTTTACA CCAAAGTTGA	4980
CAATCTTGAC ACTTTGTTGA CAACTCTGTT TTCATACAAT TTGGACAGTT CTCTTCCCA	5040
ATTCTTTCAA AAGTAGAATC ACAGTCTGAA CAAAGACAAG AGTCATCATT CCTCAGAAGT	5100
AAGAGACTAC TAAAAGTTAA AACAGTCTTC ATAGTCTGCC CACATAACAA GCACTTCATA	5160
GACCAGCCTC CTTATTCATC ATCTGAATTT CCTTAATCGC CTTCTTGATT GAAGCATTTA	5220
ACCCATCATG GAAGAAAAGC AAATCTCCTG TCGGTCTATC CATGCTTCGT CCAACTCGTC	5280
CACCAATCTG AATCAAACTA GACTTGGTAA ACAACGATG ATTGGCCTCT ACTACGAAAA	5340
CATCCACACA AGGGAAGGTA ACTCCGCGCT CCAAGATTGT CGTACTGATA AGTATTGTCA	5400
GTTCTCCATC TCGAAAAGCT TGTACTTGCT CTAATCGATC CTCTGTTACA GAAGATACAA	5460
AGCCAAATTT CTCATTGGA AATTGCTCCT GTAAGATTTC TGCTAACTGC TCCCCTTTCT	5520
TAATTTCTGA AGCAAAAATG AGTAACGGAT AAGCTGTCTT TCTCTGCTTC TCAATATAGG	5580
ACTTTAACTT TGGTGACAAA CGATTCTTGT CTAAGTAGCG ATTAAATCC GATAACCAAA	5640
TTGGTTTTGG AATAATCAAC GGATTCCAT GAAACCGTCT CGGTAAATTC AGTCTTTTTA	5700
GTTCTCCTAA ACGGACCTTT TTATCTAACT CATTGGTCTGA AGTCGCTGTT AAAAAGATTC	5760
TCAATCCATT CTCCTTTACA CTATTCTTGA CAGCGTGGTA AAGCATGGGA TTATCAACAT	5820
AAGGAAAAGC ATCTACTTCA TCCACTATCA GCAAATCAAA AGCTTGATAA AACTTCAATA	5880
ACTGATGGGT TGTTGCAACA ACTAGTGGTG TTCGAAAATA AGGTTCCGAT TCTCCATGTA	5940
GCAAAGCTAT CCCGCAAGAA AAATCCTGTT GCAGGCGCTT GTACAGCTCC AAACAAACAT	6000
CTATGCGAGG ACTAGCCAAA CAACTGCAC CACCCGCAAT GATCACTTTA GCCACTACTT	6060
GATAAATCAT TTCTGTCTTT CCAGCTCCTG TTACCGCATG AACTAAGGTT GGCTTTTGCT	6120
TGTCTACTAC TTGAAGCAAT CCTCTGACA CCTTCTCTTG AAAAGGAGTT AATTGGCCGC	6180
GCCATTTGAG AACATCTTGC TTTGGAAAAT CCTCCTCGCG AAAATAGTAT AAAGTTTGAT	6240
CACTTCTGAC TCGCTTCATC AGCAAGCACT CTCGACAATA GTAAGCACCG ATGGGCAAT	6300
ACCATTCTTC TAGAATAGTA CTATTACAGC GTTGACAGAA AAGTTTCCCC TTCTCCTTTC	6360

237

TCATTGCTGG AAGTTTCTCC GCCAACTGAC GTTCTTCTTC TGTTAATTCA TTCTCAGTAA	6420
ATAAACGACC GAGATAATCT AAATTTACTT TCATACTTCT TTATTCGTAA AAAC TAGCAC	6480
TTTAGATGAT TTTT TAGTAC AATTAAATCA TGGAATTTAG GACAATTAAA GAGGACGGTC	6540
AAGTCCAAGA AGAAATCAAA AAATCTCGCT TTATCTGCCA TGCCAAGCGT GTTTATAGCG	6600
AAGAAGAGGC TCGTGACTTC ATTACTGCCA TCAAAAAAGA AACTACAAA GCGACACATA	6660
ACTGCTCTGC CTTCAATTAT GGAGAACGTA GTGAAATTAA ACGTACAAGT GATGATGGTG	6720
AGCCTAGTGG TACTGCTGGT GTTCCCATGC TTGGGGTACT AGAAAAATCAC AATCTCACCA	6780
ATGTCTGTGT GGTCTGACA CGCTACTTTG GTGGTATTAA ACTAGGCGCT GGAGGACTAA	6840
TTCTGTCTTA CGCCGGCAGT GTCGCCTTAG CTGTCAAAGA AATTGGTATT ATTGAAATAA	6900
AAGAACAGGC TGGCATTGCT ATTCAAATGT CTTATGCTCA GTACCAAGAG TACAGTAACT	6960
TCCTTAAAGA ACATGGTCTC ATGGAGCTGG ATACAAACTT TACAGATCAA GTCGATACGA	7020
TGATTTATGT TGATAAGAA GAAAAAGAAA CTATTAAAGC TGCACCTGTG GAGTTTTTTA	7080
ATGGAAAAGT CACTTTAACT GACCAAGGTT TACGAGAGGT TGAAGTTCCT GTAAACTTAG	7140
TGTAAACAAT GAATAATACA GCGTTTCGTT GACATTCTCA CAACTACTTT AGCGAGCAAA	7200
ATAAAAAGAG CCGTACCAAA ATATACTAGA AAATGAAGCA ATTCAAACGA AACCTGATAT	7260
CGTTTTCTTT CACACCTATT TACTAGAATT AGCTGAACGC AATCACTTGA AAATTAATGA	7320
CTTTGATCTA TGATATATAG AAATGGTATG GATAGCGTTA TACTAAAGAT ATCTTATACA	7380
AAGAGGTATT CATATGTCTA TTTATAACAA CATTACTGAA TTAATCGGTC AAACACCGAT	7440
TGTTAAACTT AACAACATCG TGCCAGAAGG TGCTGCAGAC GTCTATATAA AGCTTGAAGC	7500
ATTTAATCCT GGTTCATCTG TAAAAGACCG TATTGCCCTT AGCATGATTG AAAAAGCTGA	7560
ACAAGATGGT ATTCTGAAAC CTGGTTCTAC TATTGTTGAA GCAACAAGTG GAAACACCGG	7620
TATTGGACTT TCATGGGTAG GTGCTGCTAA AGGGTATAAA GTCGTCATCG TTATGCCTGA	7680
AACTATGAGT GTAGAACGAC GTAAAATTAT CCAAGCTTAT GGTGCTGAAC TCGTCCTAAC	7740
TCCTGGTAGC GAGGGAATGA AAGGTGCTAT TGCTAAGGCT CAAGAAATCG CTGCTGAACG	7800
TGATGGTTTC CTTCTCTTTC AATTTGACAA TCCAGCTAAT CCAGAAGTAC ACGAAAGAAC	7860
AACAGGAGCT GAGATACTAG CTGCTTTCGG TAAAGATGGA TTAGATGCCT TTGTTGCTGG	7920
AGTAGGTA CT GGTGAACGA TTTCTGGTGT TTCTCATGCA CTCAAATCAG AAAATTCTAA	7980
CATTCAAGTT TTTGCAGTAG AAGCAGATGA ATCTGCTATT CTATCTGGTG AAAAACCTGG	8040
TCCTCACAAA ATTCAAGGTA TCTCAGCTGG ATTTATTCCT GATACACTTG ATACTAAAGC	8100

238

CTATGATGGT ATCGTTCGTG TAACATCAGA TGACGCTCTT GCACTCGGAC GTGAAATTGG	8160
TGGAAAAGAA GGCTTCCTTG TAGGGATTTC CTCAGCTGCA GCTATCTACG GAGCCATCGA	8220
GGTTGCCAAA AAATTAGGTA CAGGTAAAAA AGTCCTTGCC CTAGCACCAG ATAACGGTGA	8280
ACGTTATCTC TCTACAGCAC TTTATGAATT GTAACCGTCC AATAACGAAG TCTATTGAAA	8340
AATCTCCAGA CTAGAGAACT CACGGATAGT TCCTAATCTG GAGATTTCTT ATTTGCACTT	8400
TTCTTGTA CA ACTTTAGTCC ATGGTAAATA GGCCTCTAAA ACCTCTTTGT TTACGAGAGT	8460
TTCCACGTTT GGAAGACATT CTAGAAGATA GGATAGATAT TTCTCACTAT TTATAATGGA	8520
TTGAAATAAG ATATGAACAA ATCGATTAGA ACATGATGGT AAAGCGTAAT CCCTTGTTTC	8580
TCAGCTTTCC CAGACAAAAA AGTCCAATAG TAAGTCAGCT GACTATCACT CTCTAGCACC	8640
CTATAAGAAG TTTCATCCGC ATGAAGTAAG GGCTGAGTCA ATAGTCTCTC TCGCAAGAGG	8700
TTATAAAGGG GCTCCAAATA GTATTGACTC GTCTTGATAT GCCAATTAGA GATTTCTTA	8760
CGTGTGATTG GTAAACCCAT CCTAGCCCAA TCTTCTTCTT GGCATAATT GGGTACCTTC	8820
AGATTAAACT TCTGATGGAT GGTGTGAGCG ATAATAGAAG CTGAGCCAAA GTTATGCGCT	8880
AAAGGGGCTT TAGGAATAGG AGCTTTCACA AGCTTATCCA GATGATTATC TTTTACTCGT	8940
TATGGACAAT GCTATATGGC ATAAATCAAG TACCTTAAAG ATTCCGACTA ATATTGGCTT	9000
TGCATTATT CCTCCATACA CACCAGAGAT GAACCCCAT GAACAAGTGT GGAAGAGAT	9060
TCGTAAACGT GGATTTAAGA ATAAAGCCTT TCGAACTTTG GAAGATGTCA TACAAGGACT	9120
GCAGAAGGAG GTGATAAAGT CCATCGTTAA TCGGAGACGG ACTAGAATGC TTTTGA AAA	9180
CAGATGAGTA TAAAAAGAAA GTCCTCATTT CAATAGAAAT CACGACTTTC TGATGAATTT	9240
ATAGTAAAAT GAAATAAGAA CAGGATAGTC AAATCGATTT CTAAACAATGT TTTAGAAGCA	9300
GAGGTGTACT ATTCTAGTTT AAATCCACTA TATTTGGGGA GTGATAGAAA AGCCCTTCAT	9360
CAGCCAATCT ACTTGTTTCA GTGCGAGAGC TTTGACATCC TTTTCTGTAC TGGACCAAGT	9420
CAGTTTTCCTG TTCTCAAAGC GTTTATATAA TATCCAAAAT CCTTGACCAT CCCAGTAAAG	9480
AACTTTAAAG CGGTCTTTAC GTCCACCACA AAAGAGAAAG ACTTGATCGG AGAAAGGATC	9540
CAATTCAAAG TGGGTTTAA CTACATAGGC TAATGAGTCT ATTCCCTGCC TCATATCTGT	9600
CTTGCCACAA ACAAGGTGAA CTTGACCTAA ATCACTTAGT TGAATTATCA TAGTACAATA	9660
CCTTTCCTCC GATAATTATT TTTTATCTGG TATACTGGAA GTTGGGGAAT TAGGATAGAT	9720
ACCTTGTTAT GACGCGCTTA CTATGAATTT GAAGTATAGT CTCCTAAATG CACTTAGCCC	9780
TTATTATAGG GCTTTTGTG TTAATTATTC TAATCGAGTG AGACTGGGGA AAAACAATT	9840
TCAGGAAAAA TCTAAGCCCT ATACAAAAAA GGAAGCAATT TGCTTCCTTT CTATTATTAG	9900

239

TTATTCAAGG CTGCTGCCAT TGTAGCTGCA ACTTCAGCTT CGAAGTCGTT TGCAGCTTTC	9960
TCGATACCTT CACCAACTTC AAAGCGAGCA AACTCAACTA CCGAAGCGTT AACTGATTCA	10020
AGGTATGCTT CAACTGTCTT GCTGTCATCC ATGATGTAAA CTTGTGCAAG AAGTGTGTAA	10080
GCTTGGTCAA CTTTAGTGTT ATCAAGCATG AAGCGATCCA TTTTACCTGG AATAATTTTG	10140
TCCCAGATTT TTTCTGGTTT GCCTTCTGCA GCCAATTCAG CTTTGATGTC AGCTTCAGCT	10200
<hr/>	
TGAGCAATAA CATCATCAGT TAATTGAGCT TTTGATCCAT ACTTCAAGTG TGGAAGAGCT	10260
GGTTTATTAA CCATTGCACG GCTTTCGTTG TCTTGGTCGA TAACGTGATT CAATTGTGCC	10320
AACTCATCTT TAACGAATTG CTCATCCAAT TCTTTGTAAG AAAGAACTGT TGGTTTCATC	10380
GCTGCGATGT GCATTGACAA TTGTTTAGCA AGTGCTTCGT CTCCACCTTC AACAACTGAA	10440
ATAACACCGA TACGTCCACC GTTAIGTTGG TATGCTCCAA AGTGTGTGC GTCTGTTTTT	10500
TCAATCAATG CAAAGCGACG GAATGAGATT TTCTCTCCGA TAGTTGCTGT TGCAGATACG	10560
TATGCAGCTT CAAGAGTTTC ACCTGAAGGC ATTATCAAAG CAAGAGCTTC TTCGTTGTTA	10620
GCAGGTTTTT CTTTCAAGT GACTTTAGCT GTAGTATTTA CCAATTCAAC GAATTGAGCG	10680
TTTTTTGCAA CGAAGTCAGT TTCAGCGTTT ACTTCAATAA CTGCTGCAAC ATTACCGTTA	10740
ACATAAACAC CAGTCAAACC TTCTGCAGCA ACACGGTCAG CTTTCTTAGC TGCCTTAGCC	10800
ATACCTTTTT CACGAAGCAA TTCAATCGCT TTTTCGATGT CACCGTCTGT TTCTACAAGC	10860
GCTTTTTTAG CGTCCATAAC ACCGGCACCA GATTTTTCAC GCAACTCTTT TACAAGTTTA	10920
GCTGTAATTT CTGCCATTTT AATTCTCCTA TATTTTTTGA AAATAGGAGA GCGCGGCTAA	10980
GCCCCGCCTC CGG	10993

## (2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8411 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CGACGGGGAG GTTTGGCACC TCGATGTCGG CTCGTGCGAT CCTGGGGCTG TAGTCGGTCC	60
CAAGGGTTGG GCTGTTCCGC CATTAAAGCG GCACGCGAGC TGGGTTTACA ACGTCGTGAG	120
ACAGTTCGGT CCCTATCCGT CGCGGGCGTA GGAAATTTGA GAGGATCTGC TCCTAGTACG	180
AGAGGACCAG AGTGGACTTA CCGCTGGTGT ACCAGTTGTC TTGCCAAAGG CATCGCTGGG	240

240

TAGCTATGTA	GGGAAGGGAT	AAACGCTGAA	AGCATCTAAG	TGTGAAACCC	ACCTCAAGAT	300
GAGATTTCCC	ATGATTATAT	ATCAGTAAGA	GCCCTGAGAG	ATGATCAGGT	AGATAGGTTA	360
GAAGTGAAG	TGTGGCGACA	CATGTAGCGG	ACTAATACTA	ATAGCTCGAG	GACTTATCCA	420
AAGTAACTGA	GAATATGAAA	GCGAACGGTT	TTCTTAAATT	GAATAGATAT	TCAATTTTGA	480
GTAGGTATTA	CTCAGAGTTA	AGTGACGATA	GCCTAGGAGA	TACACCTGTA	CCCATGCCGA	540
ACACAGAAGT	TAAGCCCTAG	AACGCCGGAA	GTAGTTGGGG	GTTGCCCCCT	GTGAGATAGG	600
GAAGTCGCTT	AGCTTTAATC	CGCCATAGCT	CAGTTGGTAG	TAGCGCATGA	CTGTTAATCA	660
TGATGTCGTA	GGTTCGAGTC	CTACTGGCGG	AGTAATCGAT	AAAAGGGaAC	ACAGCTGTGT	720
TCCTCTTTTT	GTATCAATTT	GTATCACCAA	GCATTTTCAT	AAGGAAGTCT	GTTATTTCCT	780
GAGAACTTTC	TTTTTTTCCA	TGTGCAATCC	AAGTTTGGA	GACACCAAAA	AGTGCATGAG	840
TTAGATAGAT	GCTACTATAT	TCTAATTCAG	TGGTATTTAG	ATTCAGTTGC	ATAAATCGCT	900
TTTGTAATC	TGTACTAAGC	ATGATATGAA	GTTTATTTTCG	TAAGAAATTT	TGGATTTCCT	960
TAGTCCCAT	TTCAGAAAGA	AGGGCAGCCA	GAAGTGGTTC	TGACTCTAGA	TATTCAAAAA	1020
CTTCTAAAAT	AGCGTCTCTT	TTGTGATGAG	CATGTTTTTG	AAAAATATAT	TCAAATGTAT	1080
GGAATAGCTT	GCTTTGATAG	TGCTCAATCA	TATCATACTT	ATCCTTATAG	TGAGTATAGA	1140
AGCTGGAACG	ACTAATTCGG	GCTTTTTCTA	CTAATTTGAC	AGTAGAAATT	TTATCAAATG	1200
GCTGTTCCAT	CAGTAATTGT	ACCATAGCAT	TTTCAATAGT	TCGCTTTGTT	TTTAAGCGTT	1260
TGTTACTTTC	TTGCATATTT	CCTCCTTGTA	AACAAATTAG	ACTATATGTC	TAAAAATAGA	1320
TTTTTTATCT	TGTAATTTAG	ATTTTTTAAT	GTATAATCTA	TTATATCAAA	ATTTTAGACA	1380
ATATGTTTAA	AAAAGGAGAA	ACTAAGTTTA	AAGAATGGA	AGCAATTTAA	AAAAAACCAA	1440
CCTTTATTAT	TGTCATGATC	GGGATTTCTC	TTATTCCAGA	TCTGTACAAT	ATCATATTTT	1500
TGTCATCAAT	GTGGGATCCA	TATGGGCAAT	TGTCTGACTT	ACCTGTGGCA	GTTGTAAATA	1560
ATGATAAAGA	GGCTTCCTAT	AATGGTAATA	CTATGGCAAT	AGGAAAAGAC	ATGGTGCCA	1620
ATTTAAAGA	AAATAAAACC	TTGGATTTTC	ATTTTGTAGA	TGAAGAGGAA	GGAAAGAAGG	1680
GATTGGAAGA	TGGCGATTAC	TATATGGTAG	TGACTTTACC	AAGTGATTTA	TCTGAAAAAA	1740
CAACTACATT	ATCCAATATT	CAATCGACAG	CAGCTTATCA	ATCATTGACA	AGTGAGCAAC	1800
AAACTGAGAT	AAGTGATTCT	GATCTCAAAA	ATTCAACTGA	TAGTATTCAA	TGGGCTCAGT	1860
CAATTGTAGC	TTTAGTACAA	GATTACAGG	GAAGTTTAGA	AACTTACAA	AATCAATCTT	1920
CTAATCTTTC	GACTTTAAAA	AATCAATCTA	ATCAAGTATC	ACCTATTACT	TCTACTTCTT	1980
TGATAGGATT	GTCAAGTGGa	TTAACAGAGA	TACAAGGAGA	TGTTACTAGC	AAATTAGTTC	2040

241

CTGCCAGTCA GTCGATTGCA TCAGGTGTAA ACGCATATAC TACAGGTGTT GATAAAGTTT	2100
CTCAGGGCGC AAGTCAACTA AGTGAAAAAA ATGCCACCTT GACAGGTAGT TTGGATAAAC	2150
TAGTTTCAGG CTCAAACACC TTGACACAAA AATCTTCTAG ATTGACAGCA GGAGTTGGTT	2220
AATTACAATC AGGATCTGGG CAATTAGCAG ACAAATCCAG TCAGTTACTT TCAGGTGCTT	2280
CTCCATTAGA GAATAGAGCT AATAAATTGG CAGATGGATC TGGGAAACTA GCAGAAGGTG	2340
GAACAAAGTT AACTTCTGGA TTGGAAGATT TACAGACAGG ACTTGCTTCT TTAGGACAAG	2400
GACTAGGTAA TGCTAGTGAT CAACTCAAAT CAGTATCAAC AGAATCTAAA AATGCAGAGA	2460
TTTTGTCAAA TCCACTCAAT CTTTCAAAAA CAGACAATGA TCAAGTTCCT GTAAATGGAA	2520
TCGCAATAGC TCCTTATATG ATATCAGTTG CTCTTTTTTT GCAGCAATAT CAACAAATAT	2580
GATATTGCG AAATTGCCTT CAGGACGTCA TCCAGAGAGC CGTTGGGCTT GGTGAAATC	2640
TTGAGCTGAA ATAAATGGTA TTATAGCTGT TTTGGCAGGA ATTTTGGTAT ATGGAGGAGT	2700
TCAGCTTATT GGTTTAACTG CTAATCATGA GATGAGAATA TTTATTCTCA TCATCCTAAC	2760
AAGTTTAGTA TTCATGTCTA TGGTGACCAC TTTAGCAACG TGAATAGCC GTATAGGAGC	2820
TTTTTCTCA CTTATTTTGC TTTTACTACA GTTAGCATCA AGTGCAGGTA CTTATCCACT	2880
TGCTTTGACA AATGATTTCT TTAGATCTAT TAATCCCTGG TTACCAATGA GCTATTCAGT	2940
TTCCGGGATTA CGACAAACAA TCTCTATCAA CAAGTCATTT TCCTAGCTGT CATACTAGTT	3000
CTATTTACTA GTTTAGGTAT GCTAGCCTAT CAACATAAGA AAATGGAAGA AGATTAAAAA	3060
AATCGACCGA TTAAGTGGTC GATTTTATAT GCCTTAGATG ACTTTCGTCT GTGATTATAG	3120
ATTCCAAATA GTAAGAGAGA AGTAAAGGAA CAGATTGCTC CAGTAATAAA ACCATTGGGA	3180
TTAGCTTGTG TAATTTCTAT TTTCTTACCA TCTTGGTAGG CAGACCAACC TTTGTCATAA	3240
GGAATGGTGA AGAAAATAGA TGTATCTTGT TGGACATCAT ATGTAGCAAA AACCTTGTTT	3300
TTAGAAGTTG ATACTGTGAC AGGTGTCTCT TTAATTTTTT GAATTGCCTC GGTGAAAGTT	3420
TTGGTATCTA AACGATAGAA GGTAGGAGAT TCAAATGATA CTTGTGAATT TCCAGGGAAA	3480
CTAACATTGA TATTGAAAGT TTTTTTCTCT TTAGTATATC CTAGATTAAA GAAGGAGAAG	3540
ACATTATCAG TTGTAAAAGT CTTTTTTTCA CCATTTACAA GGATGTCAAC CTTCTTTTGT	3600
TTATCGTTAG AAAAGTGAAG GTTTATGAAA GAGAGATAAA CTTGGCTGTT TTCTGGAAC	3660
TCAATTTGAT ACTGGATTGC TGCATCTTCA TTTGAAGAAC TTGTGACACT AATCAAATCA	3720
TTAGTATTTT CTATTTTTTC TGTTTTTTCA TAAGGTATTG GAGAAAAATA ATCAAATTG	3780



242

ACGTTAGCAA GTTGATTAA AAATGAGGCC TGATTATCCA AGGTATCTTC ATTGAACTTG	3840
ACATCATTGT AAACAGATTG ACTCGCAACT GCAATCGGAA GAGAGTATTG ATTTTCATAT	3900
AGGGTAAGAT TATCTTTTG ATAGATATCT TTAAAGCCAT ACTTATCAAT AGGACTGTCT	3960
GAGATATTGT ACTGGATACC AAATAAACTA TCAGCCAAAA TACTATTATT TGCATATCGG	4020
AGATTGAGAT TAGTCCCAGA GGATTTAAAA CCAAGTTTAT CTAAAGTAGA GCTTGATGAA	4080
CGATTTGCGAA CAGATGAAAA TTGAGAGATT CCATTGTAGT TGAATTTTCAT ACTGTCATT	4140
CCTGTCTGAG TTTGTAGTTT TTCAGTACGA GTAAATTGAT TTCCAATATA TGTGAGAAA	4200
GATTCCATAG CTGGGATATC TCGACTATAA GCACTTCGAG AAGCAAATCC CCATTCCTTA	4260
GCAATTCCGT CCATTTGAGA TGAAGCATTT AAACCTATT CAACCAGTAT AAATAAAGAG	4320
ATTAGAAATGG CAAATAGATT CACAGATATA AACTTTTGA TAACTGCAAG GAGTAAAAGA	4380
GAATAGACAA CCAAAAAATC AAGAGTAAGC AGAATATTCA AATCTGTTAA AAAAGAATAA	4440
TGCGATTTTA GATAGATGGT AGCTAAAAAT CCTGCTACTA CAAGAAAAAG CGAAACTAAA	4500
AAATTCAGTA CTTTAAGTTC TTTCAGACGC TTTAAGACTT CTGCTGCTGT GTAAATTAAC	4560
AAGGTAGAGA AAATCCAAGC ATAGCGATGT AAAAACATGT TTGGAGTATG CATGCCTTGC	4620
CAAAATAAGT CAAGAGCTTC TATGTAAAAG CTGCAATTA GAAATGCAAA GAATATTACA	4680
TATATGAGTT TCACGTGAAA CTTAATAGAT TTCAGCGTAA AAAATAAAAT GGTCAAAATA	4740
AAGGGAAATA GTCCAACAAA AATCATTGGG ATGGCCCCAT ACTTTGTTGT GTCAAAGGAA	4800
CCAATGAATT GCTTAGCAAA GAGATCAAGA TACCAGCTAC TTTCAGTTTG AAACCTTGTA	4860
ACTTCAGTCA ATTTTCCCC ATGTGTCTGT AAATCAAATA GAGTGGGAAG AGTCATAATC	4920
AAACTAGCCA TACCAGCTAA AAAGCAGATA ACTATGAAAT CAAGAACAGA TGATTTTCGA	4980
GTCTTAAAGT CCCACGAAAT TTGACAGAGA TACCAGAAAA TAAGAAACAA TACTGTCATA	5040
TATCCAAAAT AATAATTTTG AATAAATAAG ATTGACAGAC TTGTAAAGTA CAATAGGAGT	5100
TTCTTTTCAG TTATCAGTAG ATGTAAACCA GTTATAATTA AAGGAATCAA GATAAAAACA	5160
TCTAGCCAGG TTTTATCTC TAATTGACTG ACAGTGAAAC TCATCAGAGC ATAGGAAGTA	5220
GATAAGGCTA GTTTTAAAA CTGAGGGATA GATTGAAACA ATTTATTCAA ACTAAAAAAG	5280
GTTGACAGAC CAATCAATCC AAATTTTAAG AGAGTTGTCA GATAGATAGC ATCTGGCATA	5340
TTCGTTAGAT CAAAAAGTA AACCAGAGGC GCGAGAAAAC TACCCAAGTA ATAAC TAGAT	5400
AGGGCATAGA AGTTTAGCCC TAGACCACTT GTAAAGGTGT AAAACAGATT ACTATTTCCA	5460
TGTAGGATAT TTCGTAAGGC TACATCAAAA ATAACGTATT GATGAAAGCC ATCTCCTAAT	5520
AGAGGAGACT TGTCGCTATT CCAGTAGATA CTTTGAGATA GATATACTCC AGACATAATC	5580

243

ACTACAGGAA TGATGAAAGA AATAAAATAG GTTCGATATG TTTTAAAAA TGATTTCATG	5640
TTACCTCGTA GAATGATAGA AAACTCAGTT GGTAAACCCA ACTGAGTTT GAAGTTTTAT	5700
TTAGTCTTTC CAAAGTTCTT TAACTTTTGC TTGTACTTCT GCATTTTCTA GGAATTCATC	5760
GTAGGTTTCA TCGATACGGT CAATGACGCC ATTTTATAGT AAGACAATGA TATGGTTAGC	5820
CAAAGTTTGA ATAAATTCGT GGTTCATGGCT GGCAAAGATG ATTGATTCTT TAAAGTTTTT	5880
CAATCCATCA TTCAAGCTTG AGATAGATTC CAAGTCCAAG TGATTTGTTG GATCATCAAG	5940
TACAAGGACA TTTGATTTTA AGAGCATGAG TTTTGAAAGC ATGACACGAA CTTTTCTCC	6000
CCCTGACAAG ACATTTACAG GTTTGTTAAC TTCATCTCCA GAGAAGAGCA TACGGCCGAG	6060
GAAGCCACGT AGGAAAGTAT TGTCATCTTC TTCTTTACTT GCGAATTGAC GCAACCAGTC	6120
AAGAATTGAT TCTCCTCCTG CAAAATCAGC TGAGTTATCT TTTGGTAGGT AAGATTGACT	6180
AGTTGTAACT CCCCCTTGA CAGTTCCTTC ATAGTCAATA TCTCCCATGA TTGCACGAAT	6240
TAATGCAGTC GTTTGAATAT CATTTGTCC AATAAGTGCT GTCTTATCAT CTGGACGCAA	6300
GATGAAACTA ATATTATCCA AGATAGTTTC ACCATCAATC TTTACAGTTA AATTTTCTAC	6360
TGTCAAGAGA TCATTACCAA TCTCAGTTC CGCTTTAAAG TTGATAAATG GATATTTACG	6420
ACTAGATGGC ACAATCTCTT CTAGCTCAAT CTTATCAAGC ATTCTCTTAC GTGATGTTGC	6480
CTGCCTTGAC TTAGAAGCAT TGGCAGAGAA ACGAGCAACA AATTCTTGCA ATTGTTTAAAT	6540
TTTTCTTCT GCTTTAGCAT TACCGTCTGC TAGCAATTTA GCAGCAAGCT CAGAAGATTC	6600
CTTCCAGAAG TCGTAGTTTC CGACATAGAG TTTGATTTTT CCAAAGTCAA GGTCCGCCAT	6660
GTGAGTACAA ACTTTGTTTA AGAAGTGACG GTCGTGGGAT ACTACGATAA CTGTGTTATC	6720
AAAGTCAATC AAGAACTT TACCAAGT ATGCAATTCG ATTTTCTTAC CTTAGTAGG	6780
CTCGTCCAAG AGAAGAACAT CTGGTTTACC AAAAAGTGCT TTGGCGAGGA GAACCTTTAC	6840
TTTTTCACCG TTGGCCAATT CGCTCATGTT TTGGTAGTGT AATTCTTCTG GAATGTTTAG	6900
GTTTTGAAGT AGTTGAGAGG CTTCACTCTC TCCTTCCCAA CCTCCAAGTT CGGCAAACCTC	6960
TCCTTCGAGT TCGGCAGCAC GAACCCCGTC CTCGTCTGAG AAATCTTCCT TCATGTAGAT	7020
AGCATCTTTC TCTTTCATGA TGCTATAAAG TTTTTCATTT CCCATGATAA CGACATCAAT	7080
GGCACGTTCA TCTTCGTAGT CAAAGTGATT TTGACGAAGA ACAGAGAGAC GTTCATCTGG	7140
ACCAAGAGAG ATGTGACCAG TAGTAGGTTT GATATCTCCA GCTAAAATTT TAAAAAAGGT	7200
TGATTTTCCG GCACCATTAG CACCGATTAA TCCGTAAGTA TTTCCTTCTG TAAATTTGAT	7260
ATTGACATCA TCAAAAAGTT TGCGATCACT AAAACGTAGT GAAACATCAG ATACTGTAAG	7320

244

CAATGTTTT CTCCTATATG TGTAATATAT TTATTCTACT AGAAAATACA GAAATATTCA	7380
AATTTTTATT TGTCAATTTT GTGTAAATTA TATTTACAGT ATCCTTTACA CAAATCTGTA	7440
AAAAGCAAGG CTGATTTATT TTGATAAATT ACGGTATTTT CATTAAAAAA ATGCTATAAT	7500
TGAAAGGACT ATATCGAAGG AGAACAAAAT GACTAAACCC ATTATTTTAA CAGGAGACCG	7560
TCCAACAGGA AAATTGCATA TTGGACATTA TGTGGGAAGT CTCAAAAATC GAGTATTATT	7620
ACAGGAAGAG GATAAGTATG ATATGTTTGT GTTCTTGGCT GACCAACAAG CCTTGACAGA	7680
TCATGCCAAA GATCCTCAAA CCATTGTAGA GTCTATCGGA AATGTGGCTT TGGATTATCT	7740
TGCAGTTGGA TTGGATCCAA ATAAGTCAAC TATTTTATT CAAAGCCAGA TTCCAGAGTT	7800
GGCTGAGTTG TCTATGTATT ATATGAATCT AGTTTCGTTA GCACGTTTGG AGCGAAATCC	7860
AACAGTCAAG ACAGAGATTT CTCAGAAAGG ATTTGGAGAA AGCATTCCTG CAGGATTCTT	7920
GGTCTATCCA ATCGCTCAAG CAGCTGATAT CACAGCTTTC AAGGCTAATT ATGTTCTGT	7980
TGGGACAGAT CAGAAACCAA TGATTGAGCA AACTCGTGAA ATTGTTCTGT CTTTAAACAA	8040
TGCATATAAC TGTGATGTCT TGGTAGAGCC GGAAGGTATT TATCCAGAAA ATGACAGAGC	8100
AGGGCGTTTG CCTGGTTTAG ATGGAAATGC TAAATGTCT AAATCACTAA ATAATGGTAT	8160
TTATTTAGCT GATGATGCGG ATACTTTGCG TAAAAAAGTA ATGAGTATGT ATACAGATCC	8220
AGATCATATC CGCGTTGAGG ATCCAGGTAA GATTGAGGGA AATATGGTTT TCCATTATCT	8280
AGATGTTTTT GGTCTCCAG AAGATGCTCA AGAAATTGCT GATATGAAAG AACGTTATCA	8340
ACGAGGTGGT CTTGGTGATG TGAAGACCAA GCGTTATCTA CTTGAAATAT TAGAACGTGA	8400
ACTGGGTCCG G	8411

## (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9064 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

TGCCGTACTC AAGTACAGCC TGCCTAAGT TTCCTAGTTT GCTCTTTGAT TTTCATTGAG	60
TATTAGTAAC CAAAATCCGA CCACATAGCC AGCCCCTATG AATATAGCCA TTAAAGCTAG	120
CATGGAATTT AGGAAATTAA AAACCACCGC AGATACAAAG GTTAGCACAA AAACATTAAA	180
AGCAATGGTG TCAGAAGCCA AGACTAGAAT ATAGGGTGTC AACCGATCTA AAGTTTGGGA	240
ATCTAGGAAA AATAAGTGT TATACATGAT GACCTCCTCT ATGGCTGAAA AGCAAGCCTT	300

245

TTGTTTTTTT	ACCCCAAGAC	CCTATGTAGA	AAAGTGAGCA	AAAACGGGAA	GGTCGCTACA	360
ATATTATTGA	TCACATGCAC	CGCATAGGAT	GGATAAATGC	TCTTGCTATA	GCGGGTCAAA	420
CCAGCAAAGA	TGATTCCAAC	TGTTGCAAAG	ACGAAGATAT	CTAACAGACT	AGGCAGGCTT	480
GAAAAATGAG	GGAGAGCAAA	TAAAATAGAA	GGAAGAAGCA	AATCAAGACC	AAATCGCGAA	540
TGCTTAAAGA	AAGCATGTTG	CAGTAATCCT	CTATAAATCA	ATTCTTCCAT	CAGTGAAGCC	600
AGAAAGAACA	GGGCTATATA	AATACCTAGC	TCTGCAAAGT	TAGTCCCACT	ATAACCAATC	660
AATACAGCCC	AACCTTCCGC	AGTTGACTGA	ACATGTTTAG	CTGTCTGAAC	GTTAAAAGAG	720
ATCTGGAACA	CTAGCACTAA	TACTGTCAAA	ATCGAATACC	AAAGCCATTT	TTTTCTTGGA	780
ATGCGGAAGA	GATAACCATG	GCCTGTCTTA	ACAAGAACCA	CAATCATGAC	TCCAATAAAA	840
AGTAAACTCA	AGATATTTTG	AATCCAGAAT	AAATTGCCTA	TCTGAGAAGA	AAATTGCCAA	900
TAGTTTGGGA	CGATAAGCGT	CAGCTGAGAA	AGACTAAATA	CGAAAAATAA	GTAAGAGAAG	960
ACTGCACTTA	TTTTGAATAG	AAGTTGATAC	TTTTTCATAG	AAATCCTCCC	TACTATGACC	1020
TCACCTTGTC	AGGCTCTACT	GCTGTAAGAT	TAAGAAGACA	GTTTGTTTTT	TTTAAGGCTA	1080
ACCTGACTAC	TAGATAATAG	ATACATTAAG	GCATTAAAGA	CAATGAAAAT	ATGTCCATAG	1140
AATAAAATCA	ACCTCGCATC	CAAACCAAGA	TAAAGTTTGA	TTATCAAAAA	GATGAGCAAA	1200
AGAATTTGAA	ACCATAAGGT	TTTTCCAAAA	ATAAATTTAA	AGCGATTTTG	AATATCTACT	1260
TCCTTGATTT	TTACCGCCAC	CCCTTTATTA	GCAAGAAGGA	AAACTCCTGC	TTCAAACAAA	1320
CCACTGTAAA	GAACAAGCCA	CCCAATAGAT	ACGATAGAGA	TTTGTA AAAA	TGTCCCTAAA	1380
AGAATATCCA	ACACACTACT	CAAGAAAATA	ACAAAAAATA	ATCTGTATTT	CATATTAAAT	1440
ACCTCCATTC	ATTTATTTCA	CTAACAATTT	AATAGAGCCT	TCTACTCAAA	TATCCTGTCA	1500
GAAAAGGATA	GAAAGCTACT	TTTTATAATA	CTTCAAGCCC	CACATGAGCA	GAAGCGTGAT	1560
AAACAAGCAG	AGAATACACC	TATATAAGCG	ATTAGTTGTT	GATAGAATTC	TGTTTCTGAA	1620
ATACCTCTAT	ACAAACAAAT	GACAAACATA	AAATCTGCCA	AGCCGATAAA	CATAAGTTGA	1680
TTGGTTCTAG	GACTAACCAA	ATCATCATTT	ACTTATATTT	AAGAGTATCT	CTTTTATTTT	1740
AATGTATGTT	AGCACTGAAA	AGCAAGACAG	GCCAATAATA	TTTAAATGA	ACAGTAACGG	1800
GGTTAAGTCT	CTAAAAAAT	TATCTACTGA	CACTACAAGA	AATACTATAC	ATATTATAGT	1860
CGAAACTATC	TTTTTCTTAT	CCATAATTAT	TTACTCCTTT	CCTAACAAAT	CCAGCTTATC	1920
AATCAAGAGC	GATTTTAAAC	ATAATGTAGC	AGCACCCGTT	GCAACTTTGA	CAAGTTTAGT	1980
ATATCATTGT	TTTTTAAAT	TTTTCATCCA	AATCTTGAAT	TGTCATCGAA	ACATCTTGAA	2040

246

TTGTTAAAAA ATTTAAAAAG TAAGCATTA	AAACATACTT TCCTCTTTAT ATTGTATTGA	2100
TACCAACTTG TTTGTAGACT TTTTCATCCTG	CTATCACATA TCATTTTGAC AGGCGAAACA	2160
ATATTAAAGA AACTCCCTG TAAATTAAGC	TAGCAAATAC AGGGGAGAAA TTTATTTTTT	2220
AGAGAGTACT ATCCGTATCC TTTTGGGAAG	ATTTTGAAAA TATTTTCTA ATTAAGTCAT	2280
CCATATAAGG ACCAAATATA CCAACTACTA	AACCAATAAT AAAACTTTTA AAATCCATAA	2340
TTACCACCAA CATATTGCTG CATAGGCTAC	ACCTCCAAGT ATAGCTCCAC CTGCAGCACC	2400
AGTTACACCT ATTCCTATAG CAAATGGTCC	CAATAGAAAT GTCAAACCGT TGTTCACAC	2460
CCATCAATTG CGCCATATGC AACCCCTGCT	GCACAATAA TTTTCTTCC CCAATCAATA	2520
TCTCCACCTT CAACGCAAGC AAGCATTTCA	TTATCCATAA CTGCAAATTG TGACATCAAT	2580
TTTGTATCCA TATAGTGTAT CACTTTTCAG	TTACGGAACA AGTTTAATAT AAAAATTATC	2640
AAAAAACAT AGGCAATAA GAGAAAAATT	AATTTATCAT AGATTAGAAA TAATATGACA	2700
AAACAATTCA ATGATGTAA TTCAATAGTC	TTTTGTTTTT TATCGGAGAT ACTTATGGAT	2760
AGATAAATAA GATAGGTTTG AAAAGCGAAG	AGAATAATAA AGAATATAGC CTTCAATAAA	2820
TTTAGCTTTC ATTTTATGA TGTAGCGGTA	TAGGCTAAAT ATCCACAAAC CACTGCTCCT	2880
CCAATTCTC CTATTGCAGC GCCCCATGGT	CCTAGAAGTC TCCCATATTT CACTCCACCC	2940
GCTGCACAAC CTAAAGCAGC AACTACAGCT	GCTCCTCCGG AATTACCTCC ATAAACCTCA	3000
CTCAGCATTG TTTCAATTAT ATTACAATAA	GTATTCATAC AAGTCTCCTT TTATTAAAA	3060
CCACCCGTTG CCCCTGTTAC TCCTGCCCAA	AGATCCACAC CAAATTTAGC TCCTATGTAT	3120
CCACATGCTC CCATAAATGG TGCTCCAACA	CCACTCGCAG CACAAATAGC TGTCCCTAGC	3180
CCCCAGCCAC CAAAAGCAGC ACCACCACCT	TCTAAGACAT TAGTTTGCCA ATTATTCTTG	3240
CCTCCTTCAA TACTAGATAA CATAGTTATA	TCCATTTCAT GAAATTGTTT CATAATTTTT	3300
GTATCCATGA CAAATACTCT TTTTATTTTT	TAATTTTTGT CTTGTTGTAA CTTTGACAA	3360
TTTAGTATAT CATCGTTTTT TAAAATTTTT	CATCCAGATT TTGAATAGTC ATCGAAACGT	3420
CTTGAATTGC AAAAATTACA TTAGACTTCC	TGCAAACTA GAATCCTAGT TCATGATTGA	3480
TAATACCAGC ACTCAAATTC ATTCGTAATC	CGAAGCGTTT ACGATGACTT CGATAGGTTG	3540
TTGAAAACAT TTAAACGTT TTTACTTTGG	CAAAGATGTT CTCAACCTTG CTTCTCTCCT	3600
TAGATAGCGC ATGGTTACAG GCTTTATCTT	CAACTGTTAG CGGTTTGAGT TTGCTGGATT	3660
TACGTGAAGT TTGTGCTTGA GGATATATCT	TCATGAGCCC TTGATAACCA CTGTCAGCCA	3720
AGATTTTACC AGCTTGTCGG ATATTTCTGC	GACTCATTTT GAACAACCTC ATATCATGAC	3780
AATAGTTCAC AGTGATATCC AAAGAAACAA	TTCTCCCTTG ACTTGTGACA ATCGCTTGAG	3840

TCTTCATAGC GTGAAATTTC TTTTACCAG AATCATTCGC TAATTCTTTT TTTAGGGCGA	3900
TTGATTTTAA CTTCGGTCGC ATCAATCATT ACCGTGTCCT CAGAACTGAG AGGAGTTCTT	3960
GAAATCGTAA CACCACCTTG AACAAAGAGTT ACTTCAACCC ATTGGCTCCG ACGGAGTAAG	4020
TTGCTTTCGT GAACACCAAA ATCAGCCGCA ATTTCTTCAT AAGTGGCGTA TTCTCGCACA	4080
TATTGAAGAG TGGCCATAAG AAGGTCTTCT AGGCTTAATT TAGGTTTTCG TCCACCTTTT	4140
GCCTGTTTAA GTTGATAAGC TGTTTTAAAT ACAGCTAGCA TCTCTTCAA AGTCGTGCGC	4200
TGAACACCAA CAAGACGCTT AAATCGTGCA TCAGTTAGTT GTTTACTTGC TTCATAATTC	4260
ATAGAACTAT AGTAAATGA AATAAGAACA GGATAAATCG ATCAGGACAG TCAAATCGAT	4320
TTCTAACAA GTTTTAGAAG TAGAGGCGTA CTATTCTAGT TTCAATCTAC TATACTATAC	4380
CATATTTTGT TTCGCAGGGA ATCTATTATA AAAGGGTAAG TATTGCAAAA AACTTACCC	4440
TTTTCTTTTA TACTTCATTA AGCTCTACTT TTTATAATAC TTCAAGCCCC ACATGAGCAG	4500
AAGCATGATG ATTAAGCAGA GAACAGCGCC AATATAAGCG ATTATTTGTT GGTAGGATTC	4560
TCCTGCTGTG ATACCTCTAT ACAAACAAAT AATAGACATA AAACCTGTCA AGCCGATGAA	4620
CATAAGTTGA TTGGTTCTAG GACTAACCAA ATCATCATCT TCAAACCTCTC TTATCCTCAT	4680
TTCCCTAGTG AGATAAACAG TAACCAAAAT AGAAGCCAAG TTAATAACTA CTAAGAAGAA	4740
TTGGAAACT ACGGAAAAAT TTAAGAACTG ACGAGATAGA AATAGATAAG TAGAAACAAG	4800
CAAGGGCAAC TGACCTAAGA ACAATCTCGC AAGGAAGATG TTCCGTTTTT TAGCAAGAAA	4860
AGTTTTCAAT TCTTTCTCC TTTCTTTTAA TTGATAGCAA AATAGATCAT AACTGCAATC	4920
ACATAGGCTA TGGTATAAAA TAGCTGATAC CAAGCACTCT CCCTAAGCGG ATATAGAAAG	4980
ATGGACATGA TTAGATACAG AACGAAAATA ATCAGTATTT TTTCTTCAT AAGATTTCTT	5040
CCTAAATGTG CGATTTATCT TAGTTGAGCA AGAACATTTA CACTGCTAGT ATAGCACTTA	5100
TTTGTACCTT GGATCACTCA AATCATAAAT GGTATCAAAA ACCTCTTGAA TTGTAAAAAT	5160
TAAAAAAGCA AGCATGAAAA ACATACTTTC CTCTTTATAT TGTATTGATA CCAACTTCTT	5220
TGTAGACTTT TCATCCTGCT ATCAGATATC ATTTTGACAG GCGAAACAAT ATTAAAGAAA	5280
CTCCCTGTGA AATTAAGCTA GCAAATACAG GGGAGAAATT TATTTTTTAG AGAGTACTAT	5340
CCGTATCCTT TTTGGAAGAT TTTGAAAATA TTTTCTAAT TAAGTCATCC ATATAAGGAC	5400
CAATATACC AACTACTAAA CCAATAATAA AACTTTTAAA ATCCATAATT ACCACCAACA	5460
TGTTGCTGCA TAGGCTACAC CTCCAAGTAT AGCTCCACCC GCAGCACCAG TTGCTGCACC	5520
TTGCCATGTT CCTGTTTTAA TGCCTAGTTG AAGACCTCTT GCTGCTCCTC CTCCAACACC	5580

248

TGCTTTGGCA AAATCTCCCC AATTGCATCC GCCACCTTCA ACGCAAGCAA GCATTTTCAGT	5640
ATCCATAACA GAAAATTGTG ACATCATTTT TGTATCCATG ACAAATACTC CTTTTTTAAA	5700
AAACTAAAAT AAATCAGAAT AGAATCCTCA TAATTTTACT ATAAGTCTTA CCAACTTAGT	5760
CCCAATTTAT CACCAACCAT ACCTCCTAAG CATGTTAATC CACCCCAAT TGCACCAATG	5820
TGTGCTCCAA CAAATGCACC AGCAAGTCCA GCTACTCCTA AAGTGGCCAA ACCTGCTCCA	5880
GTTCCACCAG TTATAATTCC CGTAGTGACT CCTGTAATCA GTGCATTTTG ACAATCAGTG	5940
GAGCTATACC CCGCTTCAAC TTTCGCAAGC ATTTTCAGTAT CCATAACCTC TAACTGTGAC	6000
AACATTTTGT TATTCATGAT GAATACCTCC TTTTATTTT CAATTTGTTA CCAAAGTCTT	6060
AAATTCATA AACAAATAGA TTTTTATAG TATCTTTTG ATTTTCTTAA AAAAGTATAT	6120
ACGTCTACTA TCTTCTTAAA GGTAGCAGTA CCTATTTTT AGTCTAAGAT TTCAATAATC	6180
TTGAGTATCT AAAATATCTT AATTTCTGTTA TTCTCCTTGC AATAAAAAGT TTTACTATAC	6240
TATTTATTAA CTGTCAGAAA GCAAAAAATA TTAGTAAATA ATAGTTTATA GTTAAGTTT	6300
TTATTCCTAC CAATCCATCA ACTAAGTAAA GCATCAACGA TTACATAAAC GATTGATAAT	6360
ATAATTAAAA TTTTGTAAAC TATCTTATTC TCATCATTTCT TAGATAACTT TGATATTTTG	6420
TAAGTAAGTA AATAAGACAG TAAATTAATA GCGATAATAA TACTATATTT AAGAATCATA	6480
ATCTTACAAA GAGGACATAA TTCCTGAACC TACACAAATA AGTGTGTGCTG CTCCCCCAGT	6540
TATCGGACCA GTCGCAGCAG CTAATAGTAC TGCTCCAATA CAACCACCGA TTGCAGATCC	6600
TAAATTGCCT CTTCCTCCAC TAACTATTTT GAGTCTTCA TTATCCATAA CAGAAAAATG	6660
TTCCATCATT TTTGTATTCA TGACAAATAC TCCTTTTTTC TTTTTTTATT TTTGTCTGT	6720
TGTAACCTTG ATAAGTTTAG TATATCATCG TTTTTTAAA TTTTTCATCC AGATCTTGAA	6780
TTGTCATCGA AACGTCTTGA ATTAGCTTTT TTATTCAAG CCACCTCTAA ATGTTTAAAA	6840
AAAATAATTT CTAATCACTT TTTTACCATT CAGGAAGTTT TAATGACTAT TCAAGATTTT	6900
ATAAAATATG AACTTAGTTT TATGACATAA TAGACCTATC CACTATATGA AAGGAATTGC	6960
CAATGACTTC TTATAAACGT ACATTTGTTT CTCAAATAGA TGCGAGAGAC TGTGGTGTG	7020
CTGCCTTAGC CTCGATTGCT AAATCTATG GTTCAGATTT TTCTCTAGCT CACTTGAGAG	7080
AACTTGCAAA GACCAATAAA GAAGGGACGA CTGCTCTTGG CATTTGTAATA GCCGCTGATG	7140
AAATGGGCTT TGAACAAGA CCTGTTCAAG CAGATAAAAC GCTCTTTGAC ATGAGTGATG	7200
TCCCTATCC ATTTATCGT CACGTTAACA AAGAAGGAAA ACTCCAACAT TACTATGTTG	7260
TCTATCAAAC AAAGAAAGAC TATCTGATTA TTGGTGATCC TGACCCTTCT GTAAAAATCA	7320
CTAAAATGTC AAAAGAACGC TTTTCTATG AATGGACTGG AGTAGCTATT TTCTAGCTA	7380

249

CCAAACCCAG CTATCAACCC CATAAAGATA AAAAGAATGG TCTACTAAGC AAGCTTCCTT	7440
CCTCTGATTT TCAAACAAAA ATCTCTCATT GCTTACATTG TTCTCTCAAG CTTATTGGTC	7500
ACTATTATCA ATATAGGTGG TTCTTACTAT CTCCAAGGAA TCTTGATGA ATACATTCCA	7560
AATCAGATGA AATCAACTTT AGGAATCATC TCAGTTGGTC TGGTTATCAC CTATATCCTC	7620
CAACAAGTCA TGAGCTTCTC CAGAGATTAT CTCCTAACCG TTCTGAGTCA GAGATTAAGT	7680
ATTGATGTGA TTTTATCCTA TATTCGCCAT ATTTTGAAC TTCCCATGTC TTCTTTGCG	7740
ACACGTCGTA CAGGAGAAAT CATTTCACGA TTCACAGATG CTAACCTCTAT TATAGATGCC	7800
TTGGCTTCTA CCATTCTTTC TCTTTTCTG GATGTTCTA TTCTGATTCT TGTAGGAGGC	7860
GTCTTACTGG CACAAAACCC TAATCTCTTC CTCTTTCTC TTATTTCCAT TCCTATATAC	7920
ATGTTTCATCA TCTTTCTTT TATGAAACCT TTCGAAAAA TGAACCATGA TGTCATGCAA	7980
AGTAATTCTA TGGTTAGCTC TGCCATTATC GAAGATATCA ACGGGATTGA AACTATAAAG	8040
TCGCTCACGA GTGAAGAAAA TCGCTATCAA AATATAGACA GCCAATTGT AGATTATTTG	8100
GAAAAATCCT TTAAGCTCAG TAAATATTCT ATTTTACAAA CGAGTTTAAA GCAGGGAACA	8160
AAATTAGTTC TGAATATCCT TATCCTATGG TTTGGCGCTC AATTAGTCAT GTCAAGTAAA	8220
ATTTCTATCG GTCAGCTGAT TACCTTTAAC ACACCTTTTT CTTACTTTAC AACTCCTATG	8280
GAAAAATTA TCAACCTCCA AACCAAATC CAATCTGCGA AGGTCGCTAA TAACCGTTTG	8340
AACGAAGTCT ATCTAGTCGA ATCTGAATT CAAGTTCAAG AAAACCCTGT TCATTACAT	8400
TTTTTGATGG GCGATATTGA ATTTGATGAC CTTCTTATA AGTATGGTT TGGATGAGAT	8460
ACCTTAACAG ATATTAATCT CACGATTAAA CAAGGAGATA AGGTTAGCCT AGTTGGACTT	8520
AGTGGTTCTG GTAAAACAAC TTTAGCCAAA ATGATTGTCA ATTTCTTTGA ACCCTACAAA	8580
GGGCATATTT CCATCAATCA TCAGGATATT AAAAACATTG ATAAAAAAGT CTTGCCCGT	8640
CATATTAATT ACCTACCCCA ACAAGCCTAT ATCTTTAATG GCTCTATTTT GGAAAACTTA	8700
ACCTTGGGCG GTAATCATAT GATTAGTCAA GAAGATATTC TAAAAGCTTG TGAAGTAGCT	8760
GAAATCCGTC AAGACATTGA AAGAATGCCT ATGGGCTATC AAACCTCAGCT CTCTGATGGA	8820
GCTGGTCTAT CAGGAGGACA GAAGCAACGA ATCGCTCTCG CTCGTGCTCT TTTAACTAAA	8880
TCTCCTGTTT TAATACTAGA TGAAGCTACT AGCGGTCTTG ATGCTCTGAC TGAGAAAAAG	8940
GTTATAGATA ATCTTATGTC TCTAACTGAT AAAACCATTC TCTTTGTAGC CCATCGTCTC	9000
AGTATAGCCG AACGAACCAA CCGTGTCATT GTTCTTGACC AGGGGAAAAAT CATTGAAGTT	9060
GGTA	9064



250

## (2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 7780 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```
CTCCATTTTT TTGATTTTCAT AAATAAACAA CCTCTCTGTT AATTTTGTAT AATTATAACG      60
ATATCCAAGT TACTTGTCAA GTGTTTTTTA AATTTTATC TCAAAAATAT TTTTTCGTTTC      120
AAAAAAGGA GCCATCAGTT GATTTCAGC TCCCTTTTAT ACAGAATTAA ACTATTTTAT      180
AGTTCGACAA TCTTACCTGT TTCAAAGTAG ACAACCCATT CACAGATATT TTTAGCATAG      240
TCACCGATAC GCTCCAAGTA GGAAATAACT TGGAAATAAT CACGACCCGT AACAAATGGCT      300
TCTGGATTTT TCTTAATCTC TTCAGTCGCA AGGTCACGGA TAGTTTCAAA ATAGTGGTTA      360
ATTTGCTCAT CCATGGAGGC CACCCGGTAT GCGTCGTCAA CAGAACCATT AAGATAAAGA      420
TCAAGTGCTG CTTCCACAAC GCTTTTAACT TCACGTCCCA TTTTTTTAAAT TTCTTCCTCT      480
ACAGCTGGAA TCGGCTCTTC CCCCTTCATA CGGATGGTTG CCTGGGCAAT GGCTACAGCG      540
TGATCCCCCA TACGCTCCAC ATCTGATACA GCCTTAAGGA CAGTCAAGAC TGTACGCAAA      600
TCTTGAGAGA CTGGTTGTTG GAGTGCGATC ATTTCAAATG ATTTCTTTTC CAGTTTCACT      660
TCGTATTCAT TTAATTCTGC ATCATCTTCG ATGACCTCTT TTGCCAGGTC ACGGTCATGC      720
GTGACAAAAG CACGTACCGT ACGATTGATT TGTGAGAGCA CTTCTTGTCC CATAGCGTAG      780
AACTGGTTAT GTAATTTCTC TAAATCTTCT TCAAATTGAG ATCGTAACAT CTTTCATCTC      840
CTTATCCAAA TTTTCCTGTA ATATAGTCTT CCGTTTCCTT GTGTTGGGGA TCAAGGAACA      900
TCTGCTTGGT ATCATTAAAT TCAATCAAAT CTCCATCTAG GAAAAATCCT GTCTTATCAG      960
AGATACGTGA AGCTTGCTGC ATGGAACGGG TTACCAGAAG CATGGTGTAC TTGTCTTTTA     1020
GACCATACAA GGTTCCTCA ATTTTACCAG CTGAAATCGG ATCCAAAGCC GAAGTTGGCT     1080
CATCCAAGAG GATGATTTTA GGA CTAGTTG CCAAGACACG GGCCACGCAG ACACGCTGCT     1140
GTTGACCACC TGACAATCCA ATAGCTGAAT CATATAGACG ATCCTTGACC TCATCCCAGA     1200
TAGAGGCACC TTGCAAGGCT TTTTCTACGG CTTTATCCAG AACCTGCTTA TCCTTAATTC     1260
CATTGATACG AAGCCCGTAG ACAACATTCT CATAGATAGT CATAGGGAAA GGATTAGGTT     1320
GTTGGAAAAC CATTCCGATT TCCTTACGTA ATTCAACCGT ATCTGTACGC GGA CTGTAGA     1380
TGTTGTGACC ATTGTACACC ACGGATCCAG TTGTGGTCAC CTCTGGATTG AGATCTCCCA     1440
```

251

TGCGGTTGAG AGACTTGAGG AGGGTTGACT TCCCTGATCC AGATGGACCA ATCAAGGCTG	1500
TAATTTCTTT AGGTTGGAAA GATAGGGAAA CACTATTCAA AGCCTTCTTT TTATTATAAT	1560
AAACGGACAG GTCTGATACC TGTAATAATCG CATCTGTCAT ACGGTTTCCCT TTCTAACCA	1620
AGTGACCAGA TACATAGTCA TTGGTGGACT GTAGCTTGGC ATTTTGGAAA ATAGTTGCAG	1680
TCTTGTCATA CTCAATCAAA TCACCCAAGT AAAAGAAGCC TGTATAGTCA CTTGCACGAG	1740
CAGCCTGCTG CATATTATGC GTTACAATGA TGATGGTAAA GTTTTCTTTG AGCTCAAACA	1800
TGGTCTCTTC TAGTTGCATG GTCGCAATCG GATCCAAGGC TGAGGCTGGC TCATCCATTA	1860
AGAGGATATC TGGCTTAACA GAGATGGCAC GAGCGATACA GAGACGTTGT TGCTGACCAC	1920
CTGATAAGGT CAAGGCTGAC TTGTGGAGAT CGTCTTTAAC CTGATCCCAG AGGGCAGCCT	1980
GACGAAGGGA GGTTCCTACG ATTTTCATCTA GGACTTGCTT ATCCTTAACT CCAGCACGTT	2040
CATGCGCAAA GGTAATATTA CGGTAAATTG ACTTAGCAAA TGGATTGGGA CGTTGAAAAA	2100
CCATTCCAAT GTGTTTACGC ATTTTCATAAA CGTTGATTTT TGGACGGTTG ACATCAATTC	2160
CACGATAGAG AATCTGCCCA GTTACTTTAG CAATATCAAT AGTATCATTG ATGCGATTGA	2220
GACTGCGTAA GTAGGTAGAT TTCCCCGATC CCGACGGGCC AATCAAAGCT GTAATTTTAT	2280
TTCTTTTCAA TTGCATATCA ATCCCCTTAA TGGATTCAAT TTTACCATAG TAAACATGGA	2340
CATCCTTAGT AGAAAGGGCT ACTTTTTCTT CAGGAAAGGT AAGGATATGC TTCTCATCCC	2400
AGTTATATGT TGACATGGCT TCTCCTTTAG GCAGCGGTTA ATTTCTTGTG TAGATAGCTT	2460
CCGAACCTAC GAGCTCCAAA GTTAAAAATC AGGATAAAGA TCAGGAGCAC AGCGGCAGAA	2520
CCTGCTGATA CAATGGTTCC ATCTGGAATA GTGCCTTCAC TATTGACTTT CCAGATATGG	2580
ACAGCCAAGG TTTCTGCTTG ACGGAAGATA GAGATGGGGC TAGTCACACT GAGGATATTC	2640
CAGTTAGACC AGTCAAGAGC TGGCGCCGAT TGCCCTGCTG TATAGATCAG AGCTGCAGCT	2700
TCGCCAAAGA TACGACCAGA TGCCAAGACG ACACCCGTTA CAATACCTGG AAGCGCTTCC	2760
GGAATAACAA CATGAACCAC TGTCTCCCAG CGAGAAATCC CAAGAGCCAG ACCAGCCTCA	2820
CGTTGGGTAT GGTGAACGTG TTTCAAATA TCCTCTACAT TACGCGTCAT CTGAGGCAAG	2880
TTAAAGACTG TCAAGGCCAA GGCACCTGAA ATGATTGAAA ATCCATACTC AACTGGACTT	2940
ACAAAGATCA AGTAACCAAA GAGACCCACC ACCACTGATG GTAAAGAGGA CAAAAATTCA	3000
ATACAAGTCC GCACAAAGTT GGTAAACAGGA CCTTTTCTAG CATATTCAGC CAAGTAAATC	3060
CCAGCTCCCA TAGAAAGAGG TACAGAAATA ATCAAGGTAA TGACCAATAG GAAAAAGGAA	3120
TTGTAAAGCT GAATGCCAAT CCCACCACCT GCTTGAAAAA CAGAAGACCT TCCAGTCAAG	3180

252

AAAGACCAAG AGATATGGGG CAAGCCCCGA ACCAAGATAT AGAGAATCAA GGAAGCCAAG	3240
ATTGTCACAA TGATGCTAGC AATCGTATAG AGGACAGCTG TTGCAAGTTT ATCTAATTTT	3300
TTAGCGCGCA TAATTTTCTT TCCCTCTTTC TTTGTAATC AATTTAATCA CACTGTAA	3360
AACTAAGCTC ATCAAGAGCA GTACCAAGGC CAGTGACCAG AGAACATTAT TATTTACAGT	3420
TCCCATGACA GTGTTCCCAA TTCCCATAGT TAATATAGAA GTTAAAGTTG CAGCTGGTGT	3480
GGTCAAGGAA GTTGGGATAA CAGCTGAGTT TCCGACAACC ATCTGGATAG CTAGAGCCTC	3540
ACCAAAGGCA CGCGCCATCC CAAAGACCAC TGCAGTGAAT ATACCAGAAC GGGCCGCTT	3600
CAAGATCACA CGCCAGATAG TCTGCCAGCG AGTGGCTCCC ATAGCGAAAC TGGCTTCACG	3660
ATAATAACGA GGAACCGCAC GCAAGCTATC CGTTGTGATA AAGGTTACGG TCGGCAAAAT	3720
CATGACAAAG AGGACGGAAA TCCCTGACAA AATCCCAAAA CCAGTCCCAC CAAAGACACT	3780
GCGAACAAAG GGAACGACGA CTTGCAAGCC AATAAATCCG TACACTACTG AAGGAATCCC	3840
AACCAGGAGT TCAATAGCTG GTTGCAAAAT CTTGCCCCCT TTTGGTGATA CTTGGTTCAT	3900
AAAACTGCT GCACCAATAG CAAAGGGTGT TGCAGTAAGG GCTGAGAGAA TGGAACGAT	3960
AAAGGAACCC AAAATCATAG GAAGGGCACC AAATCTTTTA CTAGAAGGAT TCCAAGTTCC	4020
TCCCAAAAGA AAGTCAAAGA TATTCACACC ATTGACAAAG AAGGTGCGACA AGCCTTTTGT	4080
CGCTACGAAA ACCAAAATCA TGGCCACAAG GATGACTATC AAAGAAAGAC AGGCAAAGGT	4140
CAAACCTTTT CCTAATTTCT CCAGACGAGA ATTCTTTGAT GGAAGCAACA TTTTCTTAGC	4200
TAATCTTCT TGATTCATTA TTGTCTCCCT TCCAACACTG TCACAGTTCC GGCAGCATCT	4260
TTTTCAACCT TCATTTCTCT AATCGGAATA TACTTCAATC CTTTGACAAT CCCTTCTTGG	4320
GTCTCATCCG AGAGAACAAA ATTGAGAAAT TCTGCAGCCA ACTCATTGGG CTGCCCCAAT	4380
GTATACATAT GCTCATAAGA CCACAAGGGC CAATTATTGC TACTTATATT TTCTGGACTT	4440
AAGTCATAGC CATTCAACTT CATGCTTTTG ACCGAATCAT CTATATAGGT AAGAGATAAA	4500
TAAGAGATAG CTCCTGGACT TTTTGATACG ATTGATTTTA CCGCTCCATT TGAATCCTGC	4560
TCCTGACTTT GCATGGCAGA CTGACCTTCC ATAATGACAG TATCAAAGGT AGCACGAGAG	4620
CCAGAGCCGG CTGCCCCGATT GATAACAGAG ATGGGTAAGT CCTTACCACC AACCTCTTTC	4680
CAATTGGTTA CCTCACCTAT GAAGATTTGA CGAAGTTGCT CTGTCGTTAG GTTATCAACA	4740
TCAACCTCCT TATTGACAAT CAGAGCCAAG CCAGCTACCG CGACCTTGTG GTCAACAAGA	4800
GCAGAAGCAT CAATTCCTGC TTTTCTCTCA GCAAATACAT CTGAGTTTCC TATATCAACT	4860
GCCCCAGACT GAACCTGGGA CAAGCCTGTA CCAGAACCCT CCCCTTGGAC ATTGACCGTT	4920
TTTCCAACAT GGATCGTGCC AAATTCATCT GCCGCTACTT CAACCAAGGG TTGCAAGGCA	4980

253

CTTGAGCCAA CAGCCGTTAT GGATTCTCCA CGATCAATCC AGCTAGCACA GCCTACTAAA	5040
CAAGCCGTCA GCCAAAAAGC GATAAGAGAC AGAGCAAGCT TTTTCTTTT TTTCACTGTT	5100
TTTCTCCTCG AAAATAATTA TGAATACTGT GAATTTTTTA AGTAGTCTT TATGAGTTGA	5160
CGCATGAATT CTTACCAAAT TTCTGCGCAA TTGATTATTT ATATAATATA GGCTATATTA	5220
CTCTTTCCTA ACCTCCTTTT TTCATATGTG GATAAAATCT CTGTCTATC CCTTCCCCCA	5280
TTGTACCCCA TTATAGTCAT TTCGTGTCTC TTTTCCCCT TTTAATGCA AGGGAAATTA	5340
CTCTCCTTAG ATGATAATCC AAAAGCTAGA AAGGTATCTC AAACCTCTCT ACTCTCCCAG	5400
ACTAGTTTAC AACTAAAAGG AAAAGATTCT ATTTTATGAG AAATCTAGTT TACAAGCGGT	5460
AAGAACGCTA ATAACATAAC TTCTGTACT CTTTGAAAAT CTCTTCAAAC CAGTGTTTTG	5520
AGCTATCTAT GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAG TAAAACTACA	5580
TGTAATGGCA ATCAAGATAT CAAGAATCAT CCTACTAAAA AAATCCATAC TTTCACTATA	5640
ACATAGAATA AGATATTTGA CTAGCATTTT CATTTGAATC TGAGGCCTTT TGGAAAATAA	5700
TTTTTCAAAA CATTTCCAGT AACCTTTGCA AAGCCCAAGC CATTGCCTTT AACCAAACT	5760
TGGTACCAAC CATTTGGCAG ACTTTCTGCC AGCTGAACGG TTTCTCCAGC CGCATACTTG	5820
ACAAACGCTT CTTGGCCAAT TTCAACCGAC TGTTCGACCT GACTCGGTTT CAAGGCTAAA	5880
CCAAGAGCGA AACTGGGCTC AAAGCGTTTC TTCTTAAAAG TACCCAGATG CAGTCCATTG	5940
CGAGCAATCT TGAGCTTCCA TAAATCTGGC AAAAGTTCTG GCAAGAGATA AAGCTGGTCT	6000
CCAAAAATCT GCAAGATACC CGGTAGATTG ACCTTCAAAT GGTTTTGGGC AAATTCCTGC	6060
CACAAGGCAA CTTGTTCAAG GCTGAGGTTA CTCTTACTTG CCTTAAATTT AGGAGCTGGA	6120
TTGTTACCCT TAAACTGTAG ATGGGCAACA AACTGACCCT CTCCCTTAAA CTGATGAGGA	6180
TACATCCGAG CCGTTTCTGG CAGGTCAATA CCAGCTACCA TTCCATTGAT ATGCTCTACT	6240
GGCAACAAGT CAAAATCATA CTCTCCAGC AACCAATTGA CAATCTCTC GTTTTCCTCG	6300
GGTGCCCAAG TACAGGTCGA ATAAACCAGA TGACCACCTT CAGCTAACAT GGTCACTGCA	6360
TCCTCCAGAA TTTCTCTTTG CAAGCTAGCA CATTGACTCG GATAATCTAA GCTCCAATAG	6420
TCCATAGCAT CAGGTTGCTT ACGAAACATT CCTTCACCAG AGCAAGGGGC ATCAAGAACG	6480
ATTAAGTCAA AATAGCCTTT AAAGACCTTG ACCAAGCGGT CGGCAGATTC ATTGGTCACC	6540
ACGACATTG TCGCTCCAAA ACGCTCCATG TTTTCAACCA AAATCTTAGC CCGTTTGCTT	6600
GAAATTTTCA TGGAAACAAG TAGCCCCCTC CTGCTAGAT AGGCTGCCAG TTGAGTTGAT	6660
TTGCCCCCG GTGCAGCAGC CAAGTCCAAG ACCTTCATAC CAGGACTGGG TTGGGCTACT	6720

254

TGAGCCACCA TTTGAGCAGC AGGTTCTTGC GAATAAACTA AACCTGTAGC ATGCTCAGGC	6790
GATTTCCCTG AAACCTTCCC ATAGTGGCCC CAAGGGGTTT GAGTAATGGC ATCAGAAAAG	6840
GAAAGTTGCT CTTCTTTTAA GGGATTGACC CGAAAGGCCG AAACCGCTTC CTCCTCAAAA	6900
GAGGCAAGAA AATCTCTTGC CTCATCTCCT AGTATCTCTT TATATTTTTC AACAAATCCT	6960
TCTGAAATT GCATTTAAGT TCTTTTCCTT TCGTAAATAT AGGACTGAAT TTCCTCCTGC	7020
ATCTCAAGAG GCACCATCAT GACCGGCTGT CTGGTTTGAA AATCAGGAGC TTCACCAAAA	7080
AGGGTCACAA CCCGATAGCC CAGACTTTCC CCTAAAATAC TAGCTGCGGC ATAATCCCAT	7140
GGTTGCAGAT AAGTGAGATA GGTCACAAA CGCCCTGACA AAATCTTGGC AAAACTAATG	7200
CCCGCACTTC CATAGACACG AACACCAAGA ACCGCTCGGC TCAAATCAGC CAGCCCCCAT	7260
TCATTGGTTT CCAGCATACC ACTATTCCTT GCAATGAGAA AATCTCCAAG TGGTTTAGTT	7320
TTAAAAGGAG CTAGGGACCT ATCATTTAGA CAACTGGAA ATTCCCCACC ACCGTGGTAA	7380
CAATCCCCTT TGACCACATC ATAAATCAGA CCAAATGTC CCTGACCATT TTCAAAATAA	7440
GCCATCATAA CAGCAAAATC TTCCTGCTGG GCTACAAAAT TATTGGTACC ATCAATGGGA	7500
TCAATGACCC AAACCTTGCC CTCTTGAACC GAGGCTCGCA GACAACCTTC TTCAGCACAA	7560
ATCTTATCCT CAGGATAACG GGACAAAATC TCACCAACCA AGAGTTCCTG AACTTCTTTG	7620
TCCAGTCTGG TCACCAAAATC TGTGGAGAG GACTTGGTTT CAACACGCAA GTCTTCCTGC	7680
ATATGGTCAA GAATGTACTG ACCTGCTTTC TTAACAAGCT CTTTAGCAAA TTCAAATTTA	7740
CTTTCCAAGA GAAATCTTTC CTCCCCTTT TTCTTTGGGG	7780

(2) INFORMATION FOR SEQ ID NO: 19:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4820 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double -
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

GTAATGATAT AGGAACACCA GGTGACCTGA TGGGACGTCG TAAGCCTATG AACTACTAGC	60
TGCTAAAGGC TTAAAGATG GTATGGTACC ATATATCTCA AACCAATACG AAGAAGAAGC	120
CAACAAAAG GGCAAGACAA TCAATCTCTA CGGTAAAACA AGAGGTTTGG TTACAGATGA	180
CTTGGTTTTG GAAAAGGTAT TTAATAACCA ATATCATACT TGGAGTGAGT TTAAGAAAGC	240
TATGTATCAA GAACGACAAG ATCAGTTTGA TAGATTGAAC AAAGTTACTT TTAATGATAC	300
AACACAGCCT TGGCAACAT TTGCCAAGAA AACTACAAGC AGTGTAGATG AATTACAGAA	360

255

ATTAATGGAC GTTGCTGTTT GTAAGGATGC AGAACACAAT TACTACCATT GGAATAACTA	420
CAATCCAGAC ATAGATAGTG AAGTCCACAA GCTCAAGAGA GCAATCTTTA AAGCCTATCT	480
TGACCAACA AATGATTTTA GAAGTTCAAT TTTTGAGAAT AAAAAATAGT GTCTACTATT	540
AGGAAATAAA GTTTAAAAAG GTGATGAAGA ACAAACCAAG ATTCAAGCAG GAATTCCTAC	600
TGATAATGAA GTAAGTTATG ATCTTATTTA TCAGCAGGAA ACTCTTCCTG CAACAGGTTT	660
ATCAACTTCT GAGCTTACAG CTTTAGGCCT ATTAGCTGTT GGTAGTTTAG TTCTTTTGGT	720
TCATAATATG ACGGGAACAG TTTTGTGCTC CCTCTGAAAA GTCATCATTT GATGGCTTTT	780
TTCTATATAG GGTAAAAGAT AGGGTAAAAG GCTATCATCG GACAAAATAA AGAAGGCATG	840
ATATAATATA AAGTAGATTT CTATGTCATA AAACAAGAAC TGTGGGACA TCATTCAATT	900
GAAACTCTC TATGTTCAAA CAATAGTAAA ATAAATAGG GGATCTAAAT CCTTGCTATG	960
AAAGGAAAAA ACTCAATGGC TACTATTCAA TGGTTTCCTG GTCACATGTC TAAAGCTCGT	1020
CGACAGGTGC AGGAGAATTT AAAATTTGTT GATTTTGTGA CGATTTTAGT AGATGCACGC	1080
TTGCCTCTAT CTAGTCAAAA TCCTATGTTG ACCAAGATTG TTGGTGATAA ACCAAAACTC	1140
TTGATTTTAA ACAAGGCCGA CTTGGCTGAT CCAGCAATGA CCAAGGAATG GCGTCAGTAT	1200
TTTGAATCAC AAGGAATCCA GACGCTAGCT ATCAACTCCA AAGAGCAAGT GACTGTAAAA	1260
GTTGTAACAG ATGCGGCCAA GAAGCTCATG GCTGATAAGA TTGCTCGCCA GAAAGAACGT	1320
GGGATTCAGA TTGAAACCTT GCGTACTATG ATTATCGGGA TTCCAAACGC TGGTAAATCA	1380
ACTCTGATGA ACCGTTTGGC TGGTAAAAAG ATTGCTGTTG TTGGAAACAA GCCAGGGGTC	1440
ACAAAAGGTC AACAATGGCT TAAAACCAAT AAAGACCTGG AAATCTTGGA TACACCGGGG	1500
ATTCTCTGGC CTAAGTTTGA GGATGAACT GTTGCACTTA AGTTGGCATT GACTGGAGCT	1560
ATCAAAGACC AGTTGCTTCC TATGGATGAG GTTACCATTT TTGGTATCAA TTATTTCAAA	1620
GAACATTATC CAGAAAAGCT GGCTGAACGC TTCAAACAAA TGAAAATTGA AGAAGAAGCG	1680
CCTGTGATTA TTATGGATAT GACCCGCGCC CTCGGTTTCC GTGATGACTA TGACCGTTTT	1740
TACAGTCTCT TCGTGAAGGA AGTCCGTGAT GGCAAACCTCG GTAACATAC CTTAGATACA	1800
TTGGAAGACC TCGATGGCAA CGATTAAAGA AATCAAAGAA TTCCTTGTA CAGTCAAGGA	1860
GTTAGAAAGC CCTATTTTTT TAGAGCTTGA AAAGGATAAT CGCTCAGGAG TTCAAAGGA	1920
AATCAGCAAG CGTAAAAGAG CCATTCAAGC TGAATTAGAT GAAAAATTGC GCTTGAATC	1980
CATGCTTTCT TATGAAAAAG AACTTTATAA GCAAGGATTG ACCTTAATTG CAGGTATTGA	2040
TGAGGTTGGT CGTGGTCCTC TTGCTGGTCC TGTAGTCGCT GCGGCCGTTA TTTTATCTAA	2100

256

AAATTGTAAG ATTAAAGGTC TCAACGACAG CAAGAAAATT CCTAAAAAGA AACATCTGGA	2160
GATTTTCCAA GCCGTTCAAG ACCAAGCCTT GTCGATTGGA ATTGGTATCA TAGATAATCA	2220
GGTCATCGAC CAAGTCAACA TCTATGAAGC AACCAAATA GCCATGCAAG AAGCAATCTC	2280
CCAGCTCAGC CCTCAACCAG AGCACCTTTT GATTGATGCC ATGAAACTGG ACTTGCCCAT	2340
TTCACAAACC TCCATTATCA AAGGAGATGC CAACTCCCTC TCTATCGCAG CAGCATCTAT	2400
AGTAGCCAAG GTAACACGTG ATGAATTGCT GAAAGAATAC GATCAGCAGT TCCCTGGCTA	2460
TGATTTGCGT ACTAATGCAG GATATGGCAC AGCTAAACAT CTGGAAGGCC TCACAAAAC	2520
AGGAGTTACC CCAATTCACC GAACCAGCTT TGAACCCGTT AAATCACTGG TTTTAGGTAA	2580
AAAAGAAAGT TAATTGAAAG GAAATAACAT GGAGGAACAG TCGGAAATAG TCCGTTCTAA	2640
GAAAGAATTC GCCTTTGCAT CCAGCACTAT ACTATCCCAA GTTGGTCGAG GAATCATGT	2700
CGGCCTCATC GTTGGAAATTA TCGTCGGATC CTTTCGTTTC TTAATTGAAA AGGGCTTCCA	2760
CCTGATACAA GGAGTTTATC AAGATCAAGG GTACTTAGTG CGCAATCTTT TTGTACTGGT	2820
TTTGTTTTAT AACTCATCT GTTGGCTCAG TGCCAAATA ACACGGTCAG AAAAAGATAT	2880
TAAAGGCTCA GGAATTCCTC AAGTCGAAGC CGAACTGAAA GGCCTCATGT CCCTCAACTG	2940
GTGGGGCATT CTTTGGAAAA AATATGTGCT AGGTATTCTT GCTATTGCCA GTGGACTCAT	3000
GCTGGGTCGA GAGGGACCCA GCATTCAACT TGGAGCAGTT GGTGGTAAAG GAATTGCCAA	3060
GTGGCTCAAA TCCAGTCCAG TAGAGGAACG TTCCTTGATT GCCAGTGGAG CTGCAGCAGG	3120
TTTAGCCGCA GCCTTTAATG CTCCTATTGC AGCACTTCTC TTTGTTGTAG AAGAAGTCTA	3180
TCACCATTTT TCGCGCTTTT TCTGGGTCTC AACTCTAGCA GCCAGCATCG TAGCAAAC	3240
TGTGTCTCTA CTCATGTTTG GTTTGACACC AGTATTGGAT ATGCCAGATA ACATTCTCTC	3300
CATGACCCTA GATCAGTATT GGATATATCT CGTCATGGGA ATTTTCCTTG GATTTTCAGG	3360
TTTTCTCTAT GAGAAAGCTG TATTAAACGT TGAAGAGTT TATGACTTGA TTGGTCAAAA	3420
AATCCATTG GATAGGGCTT ATTATCCCAT CTTGGCTTTT ATCCTTATCA TACCAGTCGG	3480
AATCTTCTTA CCTCAAATCA TTGGTGGCGG AAATCAGCTT GTCCTTTCTT TAACTGAACA	3540
AAATTTTAGT TTCCAAGTTT TATTAGCTTA CTTTTAATC CGCTTTATTT GGAGTATGAT	3600
TAGCTATGGA AGTGGACTGC CAGGAGGAAT TTTCTCCCC ATTTTAGCTC TTGGTTCTTT	3660
GCTTGGTGCC TTAGTTGGTG TTATCTGTGT CAATCTTGGG CTTGTCAGTC AAGAGCAATT	3720
CCCTATATTT GTCAATTCTAG GAATCAGTGG CTATTTTGGG GCCATATCAA AAGCTCCCTT	3780
AAACCGCTATG ATCCTCGTAA CTGAGATGGT AGGAGATATT CGCAACCTTA TGCCACTTGG	3840
TCTTGTCACT CTGTGTTCTT ATATTATCAT GGATTTGCTC AAAGGTACGC CAGTCTATGA	3900

257

AGCCATGCTG GAAAAAATGC TTCCAGAAGA AGTATCTAGC GAAGGAGAAG TTACACTTAT	3960
CGAAATACCA GTTCTGATA AAATTGCTGG GAAACAAGTT CATGAACTCA ACTTACCACA	4020
CAACGTCCTC ATCACAACTC AAGTCCATAA TGGCAAGAGC CAAACAGTTA ACGGCTCAAC	4080
CAGAAATGTAT CTGGGTGATA TGATTCACCT GGTATTCCA AAAAGTGAAA TTGGAAAAGT	4140
CAAAGATTTG TTGTTGTAGT ATGAGTATTT ACATAATTTA TGTTATGTAA ATGATCAGTT	4200
-----	
TGATTTATTT AGAAAACCGA TTCTCAGGAA TGAGATCGGT TATTTTTTAC TGATGAGGAA	4260
TTTTACATAT AAATAATTGA ACTTTATTAA AAATAAGACT ATAATTAAGT TAGAAATGAT	4320
AAAGTATAAA GCTAGAAAGG AGTTTACTGT ATCAAATCTG TACAGTAAGA TTAATCAT	4380
GAAAAAGAAA ACAATAGCAA TTATATAGAG AAATGAAATA GAAATAGGAT AAAACAATCA	4440
GGACAATCAA ATCAATTTCT AGCAATGTTT TAGAAGTCCA GATGTACTAT TCTAGTTTCA	4500
ATCTATTATA CAATGTGTTT TGTATCTCAT AGCTCCTTAT ATAGCTCTTC AGTTATGTAG	4560
TATTAACAGA AGTTTAGTGG GTGAGATTTT TATTATTTTC CTTATTCTGT TTTGTTGTA	4620
GGTCTAAGTC TTTTATCAC TTTGAAAAAC TCCTATAACA TCTTCCGAA AAATAAT	4680
TTTCTTGAAA AATATACAAG TCTATGCTAT ACTACTAGTA TACTTACTTA TGGAGAAAAT	4740
ACATGAAACG TGAGATTTTA CTGGAACGAA TCGACAACT AAAACAACCTC ATGCCCTGGT	4800
AAGTTCTGGA ATACTACCAA	4820

## (2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21338 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

CTACGACATC ATGATTAACA GTCATGCGCT ACTACCAACT GAGCTATGGC GGATAAAATA	60
GTCCGTACGG GATTCGAACC CGTGTTACCG CCGTGAAAAG GCGGTGTCTT AACCCCTTGA	120
CCAACGGACC TTCTATCTGT AGCAGATATA ACCATTATAT CAATTTCTTG CTAATTGTCA	180
ATCACTTTTG AGATTTTTC TCTAAAATAT CTTTAATTT TCTAATTTT AATCTTGAAA	240
TAGGACAACG ATGGTCTTCA TAGAAAACAA TTTCTAAGTT TTTTCGATCA ATTTCTCTGA	300
TATTACCTAT ATTACCAAA AATGACTTGT GAGGAGAATA AAATCGCTGA GTATGTTTGT	360
CCTTTTCCTG AATATCTGTC ATGGTACCAT AAAACTCTTT TGCAAAATTC TTACCAATAA	420



258

TGCGCAATTT ATGAGATACC CCTGTTGTTT CAATATACAA AATATCATGG TAAGGAATTT	480
TTAAATCATT TCCCTTGTA TGTAGTCGA AATAATCTAC AACATCTTCA TTTTCAAGTA	540
ACATACTCTT CGTGTAGAAG ATATTTTGCT CAATTCTCTT CTAAACATC TCATCATTTGA	600
TATCCTTATC AACAAAATCT AGGGCTGATA CCTGGTATTT ATAGGTTAGA GTCGCAAACT	660
CTGATCGACT AGTGATAAAG ACGATAATAG CGTAAGGATT GTAATGACGA ATGAGCTGAG	720
CCACTTCAAA TCCCTTTTTC TCAATTCAT GAATATCGAT ATCTAGGAAA TAAAGCTGAT	780
TTACTTCATC ATTTTCAATG TATTCTTCAA ATTCACGGAC TTTTCCGTT GTCTTGTATG	840
ATATTGGAAT ATTCGATTCT TTCGAAATTT CATCCAATAT TCTCTCTAGT CTCACTTGAT	900
GTTCAATAAC ATCTTCTAAA ATTAAAACTT TCATTCAAAT TCCCTCTTAA ATCTAATGAT	960
TTGTCTAAAT GTACTGCCTT CCATCTCTGT TTCTAAAATA ATATTGTTGT ACTTATCTAG	1020
TAGTTCTTTC ACATTATTTA ATCCGACTCC GCGATTTCTT CCCTTAGTGG AGAATCCTAA	1080
GGCAAATAGA TCTCTGAAG GAGTCATCGT CATTTTACAT GAATTCTGAA TCACAATAAC	1140
TGTTTCAGTT TCCATCTTAA TAACTGCTAC TTCCATCTGC TTTTATAGC TATCAGCCGA	1200
TCCTTCGACA GCATTATTCA ATAAAACGCT CATGATACGA ACCAAATCCA ATAGTTCAAT	1260
TGGAAGCTTG GTAATCGTAT CTTTACTTC CAGTGTAAC TCTACACCAT TATTCGAGC	1320
ATAGACAATT GACTGAGCAA CCAAACCTCG TAAAGCTGAG TCTTCTATGT TGTTCAAATC	1380
AAAGTAAGTG TACTTATCTG AACGCAATTT ATGATTTGCT TTGACTAAAA CTTCATTGTA	1440
AATTCGTCA ATTTCTGTA AATTACCACT GTCAATTGCC ATCTGCATGC TGACAAGCAT	1500
TCCAGCATAA TCATGTCGAA AACCACGGAT TTCATTATAC AGACCAACAA TTTATCTGT	1560
GTAATCTGT AAATGTTTCT GTTCAAATTT CTCTGCTTC AAAGCAATCT CTTTCTCCAT	1620
TTGAACTTTA TGAGAATTCA TTGCAAAGAA GGTCAAAGG AGAGAGATAA AGACAATAGA	1680
TGACAAAATA CTTCCAAAAC TATTCAAATG TTTAATCGTA CTTACCATAT CTGAAACGAA	1740
AGATACAATA TGTAGCAATA GTAAAGCAAA AAATACTTTT TTCAAGAAAG GATAAAGGTA	1800
GTCCTTGTC AATAGGCTA GTTCCAAATG GAAATAGTAA ATGATTTTTA ATGTAACAAA	1860
ATAGGTTAAC ACCGTCACAA CGAAAAAGAA TGGGAAATGA TATTGTAAAA CAAAATTATC	1920
TCCTGTTATA GAGGAGAAAA TTACGGACAG AAAGTTATGA GTGCTCTCAT ATAAAAGAGA	1980
TAGTAGTAAA CTTAGGAATA GTCCTCTATC CCTCTCATAC TGTTTCATCC ATCGAAAATA	2040
GGAATATAAG CCCAAAGGAA ATAAAAATCT TTCAATCCCT ATTTTATCTA AATATAGAAG	2100
ATAAAAGGAA AATTCAAGTA CTATTTCAGT TAGTAATGTA TAAGCACCAA AAACGTATAA	2160
TTCTTTTCTA TTTATTCGAC CTTTACAAAT TAAACGGTAA CTGTGACTAA TAATTAAAAA	2220

ATGAACAATA ACTGTCCCAA ATCCAAGTAA ATCCATTACT CTTTCTCCTT ATTCATTAC 2280  
TTTTTTCGTA GGAAAAGAAA ATCAAGGATG ATTCTTGAAA TCCTCATCTC CCCACCTTTA 2340  
ATCTTTTGTA AGTCTTTTTC CTTCAAAGCT ACAAAGTGT CCAATTTAAC TGTGTTTTTC 2400  
ATAATAAAAT CTCCTAAAAT GTTTTCTCTT GTAAGCTAAC TTACAAAAAC CATTATACAA 2460  
AATGGAATTT CGTTTGTAGAT AAAATCTCT CAAGTGTCT TTTTCTCTCC CAAAGTGTAC 2520  
TTTTTTAAGA AAAAGCCCG GAAAATCCC AGCTTTGCTA TTATATTGAT CCCAGCAGGA 2580  
TTCGAACCTG CGACCGTTCG CTTAGAAGGC GAATGCTCTA TCCAGCTGAG CTATGAGACC 2640  
TAATACAATT ATTCTACCAA AAATTCATT AAAAGTCAAT TTTCTATTTA TGGTAGGGGA 2700  
ATCCCTGCTG AATCGTAAAA GCGCGATAGA TTTGTTCAAC AAGAACTAGT CTCATTAAC 2760  
GATGGGGTAA GGTTAGGCGA CCAAACTGA CAGAAAGATT GGCTCTATTT TTTACAGATG 2820  
ATGATAATCC TAAACTTCCC CCAATAATAA AAGTAAGAGT AGAAAATCCT TTTATAGAAG 2880  
TTTCTCTFAA CTGCTTACTA AATTCTTCTG AGAAGAAAGT TTTCCCTTCA ATGGCTAACA 2940  
CAATAACGAA ATCACGGTCA GCAATTTTTG ATAAATTTCT CTGACCTTCT ATTTCTAAAA 3000  
TCTTTTGATT TTCTGATTCA CTGGCCTTAT CTGGTGTGTT TTCATCTGAT AACTCAATCA 3060  
TTTCAAACCT AGCAAATCTA GAAATTCGTT TTGAATACTC TGCGATACCA TCTTTTAAAT 3120  
ACTTTTCTTT CAGTTTCCCA ACTGTTACAA CTTTAATTTT CATGACTCTA TTCTAACATA 3180  
TTCTCTATTT TTTCACATCT TATTCACAAA ATAAAAATA GATTTCATT AAGAAAATCA 3240  
CAATTTCAAA AGAGTTATCC ACAGTTTGTG TAAACTTTT GTGTTTAAGT TATAATTAAG 3300  
CTAGTCAGTT TATACTTTCA GTAATTCAAA CATATGGAGG CAAATATGAA ACATCTAAAA 3360  
ACATTTTACA AAAAATGGTT TCAATTATTA GTCGTTATCG TCATTAGCTT TTTTAGTGGA 3420  
GCCTTGGGTA GTTTTCAAT AACTCAACTA ACTCAAAAAA GTAGTGTAAC CAACTCTAAC 3480  
AACAATAGTA CTATTACACA AACTGCCTAT AAGAACGAAA ATTCAACAAC ACAGGCTGTT 3540  
AACAAAGTAA AAGATGCTGT TGTTCCTGTT ATTACTTATT CGGCAAACAG ACAAATAGC 3600  
GTATTTGGCA ATGATGATAC TGACACAGAT TCTCAGCGAA TCTCTAGTGA AGGATCTGGA 3660  
GTTATTTATA AAAAGAATGA TAAAGAAGCT TACATCGTCA CCAACAATCA CGTTATTAAT 3720  
GGCGCCAGCA AAGTAGATAT TCGATTGTCA GATGGGACTA AAGTACCTGG AGAAATTGTC 3780  
GGAGCTGACA CTTCTCTGTA TATTGCTGTC GTCAAAATCT CTTCAGAAAA AGTGACAACA 3840  
GTAGCTGACT TTGGTGATTC TAGTAAGTTA ACTGTAGGAG AACTGCTAT TGCCATCGGT 3900  
AGCCCGTTAG GTTCTGAATA TGCAAATACT GTCACTCAAG GTATCGTATC CAGTCTCAAT 3960

260

AGAAATGTAT CCTTAAATC GGAAGATGGA CAAGCTATTT CTACAAAAGC CATCCAAACT	4020
GATACTGCTA TTAACCCAGG TAACTCTGGC GGCCCACTGA TCAATATTCA AGGGCAGGTT	4080
ATCGGAATTA CCTCAAGTAA AATTGCTACA AATGGAGGAA CATCTGTAGA AGGTCTTGGT	4140
TTCGCAATTC CTGCAAATGA TGCTATCAAT ATTATTGAAC AGTTAGAAAA AAACGGAAAA	4200
GTGACGCGTC CAGCTTTGGG AATCCAGATG GTTAATTTAT CTAATGTGAG TACAAGCGAC	4260
ATCAGAAGAC TCAATATTCC AAGTAATGTT ACATCTGGTG TAATTGTTCTG TTCGGTACAA	4320
AGTAATATGC CTGCCAATGG TCACCTTGAA AAATACGATG TAATTACAAA AGTAGATGAC	4380
AAAGAGATTG CTTTCATCAAC AGACTTACAA AGTGCTCTTT ACAACCATTG TATCGGAGAC	4440
ACCATTAAGA TAACCTACTA TCGTAACGGG AAAGAAGAAA CTACCTCTAT CAAACTTAAC	4500
AAGAGTTTCTG GTGATTTAGA ATCTTAATTG ACATCTATGT AAAGAAAGCT TTACATAAGA	4560
GAAAAGATGT GTTAGTGTAG AATCATGGAA AAATTTGAAA TGATTTCTAT CACAGATATA	4620
CAAAAAATC CCTATCAACC CCGAAAAGAA TTTGATAGAG AAAAAGTAGA TGAAGTAGCA	4680
CAGTCTATCA AAGAAAATGG GGTCATTCAA CCGATTATTG TTCGTCAATC TCCTGTTATT	4740
GGTTATGAAA TCCTTGACAG AGAGAGACGC TATCGGCTT CACTTTTAGC TGGTCTACGG	4800
TCTATCCCAG CTGTTGTTAA ACAGATTTC A GACCAAGAGA TGATGGTCCA GTCCATTATT	4860
GAAAATTTAC AGAGAGAAAA TTAAACCCA ATAGAAGAAG CACGCGCCTA TGAATCTCTC	4920
GTAGAGAAAG GATTCACCCA TGCTGAAATT GCAGATAAGA TGGGCAAGTC TCGTCCATAT	4980
ATCAGCAACT CCATTCTGTT ACTTTCTCTG CCAGAACAGA TTCTTTTCTG AGTAGAAAA	5040
GGCAAACTAT CACAAGCCCA TCGCGTTCC CTAGTTGGGT TAAATAACGA ACAACAAGAC	5100
TATTTCTTTC AACGGATTAT AGAAGAAGAT ATTTCTGTAA GGAAATTAGA AGCTCTTCTG	5160
ACAGAGAAAA AACAAAAGAA ACAGCAAAAA ACTAATCATT TCATACAAAA TGAAGAAAA	5220
CAGTTAAGAA AACTACTCGG ATTAGATGTA GAAATTAAAC TATCTAAAA AGACAGTGGA	5280
AAAATCATT TTTCTTTTTC AAATCAAGAA GAATATAGTA GAATTATCAA CAGCCTGAAA	5340
TAAGGCTGTT CTTTATTTT TTTATCTCAC AAGGTTATCC ACTATGTTTT TCGATAAAAA	5400
GCTTAATAAA TCAATAATTT CTCTTTTAT CCCCACCTG TGGATAAAGT TTGGTAACAT	5460
TGTGGATTAT TTTTCACAGC TTGTGAAAA TTCTTGCTAT CTATGGTAAA ATATCTCTAG	5520
TATTAACTT TTAAATAGTA AAGGAGGAGA AAGGATTGAA AGAAAAACAA TTTTGAATC	5580
GTATATTAGA ATTTGCACAA GAAAGACTGA CTCGATCCAT GTATGATTTT TATGCTATTC	5640
AAGCTGAACT CATCAAGGTA GAGGAAAATG TTGCCACTAT ATTTCTACCT CGCTCTGAAA	5700
TGGAAATGGT CTGGGAAAA CAACTAAAAG ATATTATTGT AGTAGCTGGT TTTGAAATTT	5760

261

ATGACGCTGA AATAACTCCC CACTATATTT TCACCAAACC TCAAGATACG ACTAGCTCAC	5820
AAGTTGAAGA AGCTACAAAT TTAACCTCTT ATAACCTATAG TCCAAAGTTA GTATCTATTC	5880
CTTATTCAGA TACGGGATTA AAAGAAAAGT ATACCTTTGA TAACTTTATT CAAGGGGATG	5940
GAAATGTTTG GGCTGTATCA GCCGCTTTAG CTGTCTCTGA AGATTTGGCT CTGACCTATA	6000
ACCCTCTTTT TATCTATGGA GGACCAGGCC TTGGTAAGAC TCACTTATTA AACGCTATTG	6060
GAAATGAAAT TCTAAAAAT ATTCTAATG CGCGTGTTAA ATATATCCCT GCCGAAAGCT	6120
TTATTAATGA CTTTCTTGAT CACCTAAGAC TTGGGGAAAT GGAAAAGTTT AAAAGACCT	6180
ATCGTAGTCT TGATCTTTTG TTAATCGATG ATATCCAGTC ACTCAGCGGA AAAAAGTCG	6240
CAACTCAGGA AGAATTTTTC AATACCTTTA ACGCCCTTCA TGACAAGCAA AAACAGATTG	6300
TCCTAACGAG TGATCGTAGT CCAAAACATC TAGAAGGGCT CGAGGAGAGG CTTGTCACGC	6360
GTTTTAGTTG GGGATTGACA CAACTATCA CCCCCCTGA CTTTGAAACA CGTATTGCCA	6420
TTTTACAAAG TAAGACGGAA CATTTAGGCT ACAATTCCCA AAGTGATACT CTAGAATACC	6480
TAGCTGGGCA ATTTGATTCA AATGTTGAG ATCTTGAGGG AGCCATCAAC GACATCACTT	6540
TAATTGCCAG AGTAAAAAAA ATCAAGGATA TCACTATTGA TATTGCTGCA GAAGCCATTA	6600
GAGCCCGCAA ACAAGATGTT AGCCAAATGC TCGTCATCCC AATTGATAAA ATCCAAACTG	6660
AAGTTGGTAA CTTTTATGGT GTTAGTATCA AAGAAATGAA GGGAAAGTAGA CGCCTTCAAA	6720
ATATTGTTTT GGCCCGTCAA GTAGCCATGT ATTTATCTAG AGAACTAACA GATAATAGTC	6780
TTCCAAAAAT TGGGAAGGAA TTTGGGGGAA AAGATCATAC CACAGTCATT CATGCCCATG	6840
CCAAAATAAA ATCTTTGATT GATCAAGACG ATAATTTACC TTTAGAAATT GAATCAATCA	6900
AAAAGAAAAT CAAATAATTT GTGGATAACT TTTAGTTTTT TATCTTTTTT ATCCACATTT	6960
TTTAAACAAG CTAAAAAACT TGATATGACT TGTTTAAAGG CTGTTTTCCA CAGATTTTAC	7020
AGACTCTATT ATTACTATTA TCTTTCTAAT ACTAAAAATA AATAAAGGAG AATCCATGAT	7080
TCATTTTTCA ATTAATAAAA ATTTATTTCT ACAAGCATT AATACTACTA AGAGAGCTAT	7140
TAGTTCTAAA AATGCCATTC CTATTTTATC AACAGTAAAA ATTGACGTGA CCAATGAAGG	7200
TATTACTTTA ATGGGTTCAA ATGGTCAAAT TTCAATTGAA AATTTTATTT CTCAAAAAAA	7260
TGAAGATGCT GGTGTTTAA TTACTTCTTT AGGTTGATC CTTCTTGAAG CTTCTTTCTT	7320
TATCAATGTA GTATCTAGTT TACCTGATGT AACTCTTGAT TTTAAAGAAA TTGAACAAAA	7380
TCAAATTGTT TTAACCAGTG GCAAATCAGA AATTACCCTA AAAGGAAAAG ATAGCGAACA	7440
ATATCCACGA ATCCAAGAAA TTTAGCAAG CACTCTTTTA ATACTGAAA CAAAATTACT	7500

262

CAAGAAAATT ATTAATGAAA CAGCCTTTGC TGCAAGTACA CAAGAGAGTC GTCCGATTTT	7560
AACAGGTGTC CACTTCGTAT TGAGTCAACA CAAAGAGTTA AAAACAGTTG CAACAGACTC	7620
TCATCGCCTA AGCCAGAAAA AATTGACTCT TGAAAAAAT AGTGATGATT TTGATGTCGT	7680
AATTCCCTAGC CGTTCTCTAC GCGAATTTTC AGCGGTATTT ACAGATGATA TCGAAACTGT	7740
AGAGATTTTC TTTGCCAATA ACCAAATCCT CTTTAGAAGC GAAATATTA GCTTCTATAC	7800
TCGTCTCCTA GAAGGAAACT ATCCTGATAC AGATCGCTTG ATTCCAACAG ACTTTAACAC	7860
TACTATTACT TTTAATGTGG TAAACTTACG CCAGTCAATG GAGCGTGCCC GTCTTTTATC	7920
AAGTGGGACT CAAAATGGTA CTGTGAACT TGAAATTAAG GATGGGGTTG TTAGCGCCCA	7980
TGTTCACTCT CCAGAAGTTG GTAAAGTAAA CGAAGAAATC GATACTGATC AGGTTACTGG	8040
TGAAGATTTG ACCATTAGTT TCAACCCAAC TTAGTTGATT GATTCTCTTA AAGCTTTAAA	8100
TAGCGAAAAG GTGACTATTA GCTTTATCTC AGCTGTTCTG CCATTTACTC TTGTGCCAGC	8160
AGATACTGAC GAAGACTTCA TGCAGCTCAT TACACCAGTT CGTACAAATT AAGTGAAAGA	8220
GCTTGAGCCT GGCTCGCCTC TTTTATGATA TAATCGAAAA AGAAAAGGAG AGTAGTATGT	8280
ATCAAGTTGG AAATTTTGTG GAGATGAAAA AATCACACGC TTGTACAATC AAGTCGACTG	8340
GTAAAAAGGC TAATCGTTGG GAAATTACAC GTGTAGGAGC AGATATCAAA ATAAATGTA	8400
GTAATTGTGA GCATGTTGTC ATGATGGGGC GATATGATTT TGAGCGAAAA ATGAATAAAA	8460
TTATTGACTG AGAACCTTA GTTAGAGGGT TAGCACTTTA TCCCTTTTTG TGTATAATA	8520
TTAGGGATTG AAATGAAAAA GGAGAATGAG AAATATGGCT TTGACAGCAG GTATCGTTGG	8580
TTTGCCAAAC GTTGGTAAAT CAACACTATT TAATGCAATT ACAAAGCAG GAGCAGAGGC	8640
AGCAAACCTAC CCATTTGCCA CGATTGATCC AAATGTTGGA ATGTTGGAAG TTCCAGATGA	8700
ACGCCTACAA AACTAACTG AAATGATAAC TCCTAAAAAG ACAGTTCCCA CAACATTTGA	8760
ATTTACAGAT ATTGCAGGGA TTGTAAAAGG AGCTTCAAAA GGAGAGGGGC TAGGGAATAA	8820
ATTCTTGGCC AATATTCGTG AAGTAGATGC GATTGTTTAC GTAGTTCGTG CTTTGTATGA	8880
TGAAAATGTA ATGCCGAGC AAGGACGTGA AGACGCCTTT GTAGATCCAC TTGCAGATAT	8940
TGATACCATT AATCTGGAAT TGATTCTTGC TGACTTAGAA TCAGTGAACA AACGATATGC	9000
GCGTGTAGAA AAGATGGCAC GTACGCAAAA AGATAAAGAA TCAGTAGCAG AATTCAATGT	9060
TCTTCAAAAG ATTAACCAG TCCTAGAAGA CGGGAAATCA GCTCGTACCA TTGAATTTAC	9120
AGATGAGGAA CAAAAGGTTG TCAAAGGTCT TTTCTTTTG ACGACTAAAC CAGTTCTTTA	9180
TGTAGCTAAT GTGGACGAGG ATGTGGTTTC AGAACCTGAC TCTATCGACT ATGTCAAACA	9240
AATTCGTGAA TTTGCAGCGA CAGAAAATGC TGAAGTAGTC GTTATTTCTG CCGTGCTGA	9300

263

GGAAGAAATT TCTGAATTGA ATGATGAAGA TAAAAAAGAG TTTCTTGAAG CCATTGGTTT	9360
GACAGAATCA GGTGTAGATA AGTTGACGCG TGCAGCTTAC CACTTGCTTG GATTGGGAAC	9420
TTACTTCACA GCTGGTGAAA AAGAAGTTCG CGCTTGGACT TTCAAACGTG GTATGAAGGC	9480
TCCTCAAGCA GCTGGTATTA TCCACTCAGA CTTTGAAAAA GGCTTTATTC GTGCAGTAAC	9540
CATGTCATAT GAAGATCTAG TGAAATACGG ATCTGAAAAG GCCGTAAAAG AAGCTGGACG	9600
CTTGCGTGAA GAAGGAAAAG AATATATCGT TCAAGATGGC GATATCATGG AATCCGCTT	9660
TAATGTCTAA AAATTAATAA ATGGTGTCAA TTAGGTTGGA AAAAAATTCC AACCTTTTG	9720
GCTTTTGAAA GGAATAATAA ATGACCAAAT TACTTGTAGG CTTGGGAAAT CCAGGGGATA	9780
AATATTTTGA AACAAAACAC AATGTTGGTT TTATGTTGAT TGATCAACTA GCGAAGAAAC	9840
AGAATGTCAC TTTTACACAC GATAAGATAT TTCAAGCTGA CCTAGCATCC TTTTTCCTAA	9900
ATGGAGAAAA AATTTATCTG GTTAAACCAA CGACCTTTAT GAATGAAAGT GGAAGCAG	9960
TTTCATGCTTT ATTAACCTTAC TATGGTTTGG ATATTGACGA TTTACTTATC ATTTACGATG	10020
ATCTTGACAT GGAAGTTGGG AAAATTCGTT TAAGAGCAAA AGGCTCAGCA GGTGGTCATA	10080
ATGGTATCAA GTCTATTATT CAACATATAG GAACTCAGGT CTTTAACCGT GTTAAGATTG	10140
GAATTGGAAG ACCTAAAAAT GGTATGTCAG TTGTTTCATCA TGTTTTGAGT AAGTTTGACA	10200
GGGATGATTA TATCGGTATT TTACAGTCTG TTGACAAAGT TGACGATTCT GTAAACTACT	10260
ATTTACAAGA GAAAAATTTT GAGAAAACAA TGCAGAGGTA TAACGGATAA ATGGTGACCT	10320
TATTAGATTT ATTCTCAGAA AATGATCAGA TTAATAAATG GCATCAAAAT TTAACAGATA	10380
AGAAAAGACA ACTAATACTT GGTTTATCAA CATCTACTAA GGCTCTTGCA ATTGCAAGCA	10440
GTTTAGAAAA AGAAGATAGG ATTGTGTTAT TGACGTCAAC TTATGGAGAA GCAGAAGGAC	10500
TTGTTAGTGA TCTTATTTCT ATCTTGGGTG AGGAACTCGT CTATCCATTT TTGGTAGATG	10560
ATGCTCCTAT GGTGGAGTTT TTGATGTCTT CACAGGAAAA AATTATTTCA CGGGTTGAAG	10620
CCTTGCGTTT TTTGACTGAT TCATCTAAGA AAGGGATTTT AGTTTGTAAT ATCGCAGCAA	10680
GTCGATTGAT TTTACCGTCT CCCAATGCAT TCAAAGATAG TATTGTAAAA ATCTCAGTTG	10740
GTGAAGAATA TGATCAACAC GCGTTTATCC ATCAGTTAAA GGAAAAATGGC TATCGAAAAG	10800
TTACTCAAGT ACAAACCTCAG GGCGAATTTA GTCTTCGAGG AGATATTTTA GATATTTTGT	10860
AAATATCCCA GTTAGAACCT TGTCGAATTG AGTTTCTTGG TGATGAAATT GATGGTATCA	10920
GGTCATTTGA AGTAGAAACA CAATTATCGA AAGAAAAATA GACAGAACTC ACTATCTTTC	10980
CAGCTAGTGA TATGCTTTTG AGAGAAAAGG ATTATCAACG AGGACAGTCA GCTTTAGAAA	11040

264

AACAAATTC AAAAATTTA TCACCTATT TGAAATCATA CCTAGAAGAA ATTCTTTCAA	11100
GTTTTACCA AAAACAAAGT CATGCAGACT CTCGGAAGTT TTTATCTTTG TGCTATGATA	11160
AGACATGGAC TGTCTTTGAT TATATTGAAA AAGATACTCC AATATTCTTT GATGATTATC	11220
AAAAATTGAT GAATCAGTAT GAAGTCTTTG AAAGAGACTT AGCGCAGTAC TTTACAGAAG	11280
AATTACAGAA TAGTAAAGCA TTTTCTGATA TGCAGTATTT TTCTGATATT GAACAAATCT	11340
ATAAAAAACA AAGTCCAGTG ACCTTTTTCT CTAATCTTCA AAAGGTTTA GGAAATCTCA	11400
AATTTGACAA AATTTATCAA TTCAATCAAT ATCCTATGCA GGAATTTTC AATCAGTTTT	11460
CTTTTCTAAA AGAAGAAATT GAACGATATA AAAAAATGGA TTACACCATT ATTCTGCAGT	11520
CTAGCAATTC AATGGGAAGT AAAACATTGG AGGATATGTT AGAGGAATAT CAGATTAAAT	11580
TGGATTCTAG AGATAAGACA AATATCTGTA AAGAATCTGT AAACCTAATA GAGGTAATC	11640
TCAGACATGG TTTTCATTTT GTAGATGAAA AGATTTTATT GATAACTGAA CATGAGATTT	11700
TTCAAAAGAA ATTAAAGCGT CGTTTTTCGAA GACAACATGT TTCAAATGCA GAGAGATTAA	11760
AAGATTACAA TGAACCTGAA AAAGGGGACT ATGTTGTCCA TCATATCCAT GGGATTGGTC	11820
AATATCTAGG AATTGAAACC ATTGAAATCA AGGGAATTCA TCGCGATTAT GTCAGTGTCC	11880
AATACCAAAA TGGTGATCAA ATTTCTATCC CCGTGGAAAC GATTCATCTA CTGTCCAAAT	11940
ATATTTCAAG TGATGGTAAA GCTCCAAAAC TCAATAAATT AAATGACCGT CATTTTAAAA	12000
AGGCCAAGCA AAAGGTTAAG AACCAGGTAG AGGATATAGC TGATGATTTA ATCAAATCT	12060
ACTCTGAACG TAGTCAGTTG AAGGGTTTTG CTTTCTCAGC TGATGATGAT GATCAAGATG	12120
CCTTTGATGA TGCTTTCCCT TATGTTGAAA CGGATGATCA ACTTCGTAGT ATTGAGGAAA	12180
TCAAGAGGGA TATGCAGGCT TCTCAGCCAA TGGATCGACT TTTAGTTGGG GATGTTGGTT	12240
TTGGAAGAC TGAAGTTGCT ATGCGTGCAG CCTTTAAAGC AGTCAATGAT CACAAACAGG	12300
TTGTCAATTCT AGTTCCGACG ACGGTTTTAG CGCAACAGCA CTATACGAAT TTAAAGGAAC	12360
GATTCCAAAA TTTTGCAGTT AATATTGATG TGTGAGTCG CTTTAGAAGT AAAAAAGAGC	12420
AGACTGCAAC ACTTGAAAAA TTGAAAAACG GTCAAGTCGA TATTTTGATT GGAACACATC	12480
GTGTTTTGTC AAAAGATGTT GTGTTTGCTG ATTTGGGCTT GATGATTATT GATGAGGAAC	12540
AGCGATTTGG TGTCAAGCAT AAGGAACTT TGAAGAAGT GAAGAAACAA GTGGATGTCC	12600
TAACCTTGAC CGCTACGCCA ATCCCTCGTA CCCTCCATAT GTCTATGCTG GGAATCAGAG	12660
ATTTATCTGT TATTGAACT CCGCCGACTA ATCGCTATCC TGTTCAGACC TATGTTTTGG	12720
AAAAGAATGA TAGTGTCAAT CGTGATGCTG TCTTGCGTGA AATGGAGCGT GGAGGTCAAG	12780
TTTATTATCT TTACAACAAA GTTGACACAA TTGTTTCAGAA GGTTCAGAA TTACAGGAGT	12840

265

TGATTCCGGA GGCTTCGATT GGATATGTTT ATGGTCGAAT GAGTGAAGTC CAGTTGGAAA	12900
ATACTCTATT AGACTTTATT GAGGGACAAT ACGATATCTT GGTGACGACT ACTATTATTG	12960
AGACAGGGGT GGACATTCCA AATGCTAATA CTTTATTTAT TGAAAATGCG GACCATATGG	13020
GCTTGTC AAC CTTATATCAG TTAAGAGGAA GAGTCGGTCG TAGTAATCGT ATTGCTTATG	13080
CTTATCTCAT GTATCGTCCA GAAAAATCAA TCAGTGAAGT CTCTGAAAAG AGATTAGAAG	13140
CGATTAAAGG ATTTACAGAA TTGGGCTCTG GCTTTAAGAT TGCAATGCGA GATCTTTCGA	13200
TTCTGTGGAGC AGGAAATCTT TTAGGAAAAT CCCAGTCTGG TTTCATTGAT TCTGTGGTT	13260
TTGAATTGTA TTCGCAGTTA TTAGAGGAAG CTATTGCTAA ACGAAACGGT AATGCTAACG	13320
CTAACACAAG AACCAAAGGG AATGCTGAGT TGATTTTGCA AATTGATGCC TATCTTCCTG	13380
ATACTTATAT TTCTGATCAA CGACATAAGA TTGAAATTTA CAAGAAAATT CGTCAAATTG	13440
ACAACCGTGT CAATTATGAA GAGTTACAAG AGGAGTTGAT AGACCTTTT GGAGAATACC	13500
CAGATGTAGT AGCCTATCTG TTAGAGATTG GTTTGGTCAA ATCATACTTG GACAAGGCT	13560
TTGTTCACCG TGTGAAAAGA AAAGATAATA AAATTACAAT TCAATTTGAA AAAGTCACTC	13620
AACGACTGTT TTTAGCTCAA GATTATTTTA AAGCTTTATC CGTAACGAAC TTAAGCAG	13680
GCATCGCTGA GAATAAGGGA TTAATGGAGC TTGTATTTGA TGTCCAAAAT AAGAAAGATT	13740
ATGAAATTTT AGAAGGTTTG CTGATTTTTC GAGAAAGTTT ATTAGAGATA AAAGAGTCTA	13800
AGGAAGAAAA TTCCATTTGA TATTTTCTT CTATAAAATA GATAAAAATG GTACAATAAT	13860
AAATTGAGGT AATAAGGATG AGATTAGATA AATATTTAAA AGTATCGCGA ATTATCAAGC	13920
GTCTACAGT CGCAAAGGAA GTACGAGATA AAGGTAGAAT CAAGGTTAAT GGAATCTTGG	13980
CCAAAAGTTC AACGGACTTG AAAGTTAATG ACCAAGTTGA AATTCGCTTT GGCAATAAGT	14040
TGCTGCTTGT AAAAGTACTA GAGATGAAAG ATAGTACAAA AAAAGAAGAT GCAGCAGGAA	14100
TGTATGAAAT TATCAGTGAA ACACGGGTAG AAGAAAATGT CTAAAAATAT TGTACAATTG	14160
AAATAATCTT TTATTCAAAA TGAATACCAA CGTCGTCGCT ACCTGATGAA AGAACGACAA	14220
AAACGGAATC GTTTTATGGG AGGGGTATTG ATTTTGATTA TGCTATTATT TATCTTGCCA	14280
ACTTTTAATT TAGCGCAGAG TTATCAGCAA TTAATCCAAA GACGTCAGCA ATTAGCAGAC	14340
TTGCAAACTC AGTATCAAAC TTTGAGTGAT GAAAAGGATA AGGAGACAGC ATTTGCTACC	14400
AAGTTGAAAG ATGAAGATTA TGCTGCTAAA TATACACGAG CGAAGTACTA TTATTCTAAG	14460
TCGAGGGAAA AAGTTTATAC GATTCTGAC TTGCTTCAAA GGTGATAAAA TCGAAAATTT	14520
ATTAGACGTA ATAGAGCAAT TTTTGAGTTT GTCAGATGAA AAGCTGGAAG AATTGGCTGA	14580



266

TAAAAATCAA TTATTGCGTT TACAAGAAGA AAAGGAAAGG AAGAATGCGT AAATTCCTAA	14640
TTATTTTGTT GCTACCAAGT TTTTGACCA TTTCAAAAGT CGTTAGCACA GAAAAAGAAG	14700
TCGTCTATAC TTCGAAAGAA ATTTATTACC TTTCACAATC TGACTTTGGT ATTTATTTTA	14760
GAGAAAAATT AAGTTCTCCC ATGGTTTATG GAGAGGTTCC TGTTTATGCG AATGAAGATT	14820
TAGTAGTGCA ATCTGGGAAA TTGACTCCCA AAACAAGTTT TCAAATAACC GAGTGGCGCT	14880
TAAATAAACA AGGAATTCCA GTATTTAAGC TATCAAAATCA TCAATTTATA GCTGCGGACA	14940
AACGATTTT ATATGATCAA TCAGAGGTAA CTCCAACAAT AAAAAAGTA TGGTTAGAAT	15000
CTGACTTTAA ACTGTACAAT AGTCCTTATG ATTTAAAAGA AGTGAAATCA TCCTTATCAG	15060
CTTATTCGCA AGTATCAATC GACAAGACCA TCTTTGTAGA AGGAAGAGAA TTTCTACATA	15120
TTGATCAGGC TGGATGGGTA GCTAAAGAAT CAACTTCTGA AGAAGATAAT CGGATGAGTA	15180
AAGTTCAAGA AATGTTATCT GAAAAATATC AGAAAGATTC TTTCTCTATT TATGTTAAGC	15240
AAGTACTAC TGGAAAAGAA GCTGGTATCA ATCAAGATGA AAAGATGTAT GCAGCCAGCG	15300
TTTTGAACT CTCTTATCTC TATTATACGC AAGAAAAAAT AAATGAGGGT CTTTATCAGT	15360
TAGATACGAC TGTAATAATAC GTATCTGCAG TCAATGATTT TCCAGGTTCT TATAAACAG	15420
AGGGAAGTGG TAGTCTTCCT AAAAAAGAAG ATAATAAAGA ATATTCTTTA AAGGATTTAA	15480
TTACGAAAGT ATCAAAAGAA TCTGATAATG TAGCTCATAA TCTATTGGGA TATTACATTT	15540
CAAACCAATC TGATGCCACA TTCAAATCCA AGATGTCTGC CATTTATGGGA GATGATTGGG	15600
ATCCAAAAGA AAAATTGATT TCTTCTAAGA TGGCCGGGAA GTTTATGGAA GCTATTTATA	15660
ATCAAAATCG ACTTGTGCTA GACTCTTTGA CTAAAACAGA TTTTGATAGT CAGCGAATTG	15720
CCAAAGGTGT TTCTGTTAAA GTAGCTCATA AAATGGGAGA TGGCGATGAA TTTAAGCATG	15780
ATACGGGTGT TGTCTATGCA GATTCTCCAT TTATTCTTTC TATTTTCACT AAGAATCTTG	15840
ATTATGATAC GATTTCTAAG ATAGCCAAGG ATGTTTATGA GGTTCATAAA TGAGGGAACC	15900
AGATTTTTTA AATCATTTTC TCAAGAAGGG ATATTTCAA AAGCATGCTA AGGCGGTTCT	15960
AGCTCTTTCT GGTGGATTAG ATTCCATGTT TCTATTTAAG GTATTGTCJA CTTATCAAAA	16020
AGAGTTAGAG ATTGAATTGA TTCTAGCTCA TGTGAATCAT AAGCAGAGAA TTGAATCAGA	16080
TTGGGAAGAA AAGGAATTAA GGAAGTTGGC TGCTGAAGCA GAGCTTCCTA TTTATATCAG	16140
CAATTTTCA GGAGAATTTT CAGAAGCGCG TGCACGAAAT TTTCGTTATG ATTTTTCATCA	16200
AGAGGTCATG AAAAAGACAG GTGCGACAGC TTAGTCACT GCCCACCATG CTGATGATCA	16260
GGTGGAAACG ATTTTATGCG GCTTGATTCT AGGAACTCGC TTGCGCTATC TATCAGGAAT	16320
TAAGGAGAAG CAAGTAGTCG GAGAGATAGA AATCATTCGT CCCTTCTTGC ATTTTCAGAA	16380

AAAAGACTTT	CCATCAATTT	TTCACTTTGA	AGATACATCA	AATCAGGAGA	ATCAATTATTT	16440
TCGAAATCGT	ATTGCAAATT	CTTACTTACC	AGAATTGGAA	AAAGAAAATC	CTCGATTTAG	16500
GGATGCAATC	TTAGGCATTG	GCAATGAAAT	TTTAGATTAT	GATTTGGCAA	TAGCTGAATT	16560
ATCTAACAAT	ATTAATGTGG	AAGATTTACA	GCAGTTATTT	TCTTACTCTG	AGTCTACACA	16620
AAGAGTTTTA	CTTCAAACCT	ATCTGAATCG	TTTTCCAGAT	TTGAATCTTA	CAAAAGCTCA	16680
GTTTGCTGAA	GTTCAGCAGA	TTTTAAAATC	TAAAAGCCAG	TATCGTCATC	CGATTAAAAA	16740
TGGCTATGAA	TTGATAAAAG	AGTACCAACA	GTTTCAGATT	TGTAAAATCA	GTCCGCAGgC	16800
TGATGAAAAG	GAAGATGAAC	TTGTGTTACA	CTATCAAAAT	CAGGTAGCTT	ATCAAGGATA	16860
TTTATTTTCT	TTTGGACTTC	CATTAGAAGG	TGAATTAATT	CAACAAATAC	CTGTTTCACG	16920
TGAAACATCC	ATACACATTC	GTCATCGAAA	AACAGGAGAT	GTTTTGATTA	AAAATGGGCA	16980
TAGAAAAAAA	CTCAGACGTT	TATTTATTGA	TTTGAAAATC	CCTATGGAAA	AGAGAAACTC	17040
TGCTCTTATT	ATTGAGCAAT	TTGGTGAAAT	TGTCTCAATT	TTGGGAATTG	CGACCAATAA	17100
TTTGAGTAAA	AAAACGAAAA	ATGATATAAT	GAACACTGTA	CTTTATATAG	AAAAAATAGA	17160
TAGGTAAAAA	ATGTTAGAAA	ACGATATTAA	AAAAGTCCTC	GTTTCACACG	ATGAAATTAC	17220
AGAAGCAGCT	AAAAAAGTAG	GTGCTCAATT	AACTAAAGAC	TATGCAGGAA	AAAATCCAAT	17280
CTTAGTTGGG	ATTTTAAAAG	GATCTATTCC	TTTTATGGCT	GAATTGGTCA	AACATATTGA	17340
TACACATATT	GAAATGGACT	TCATGATGGT	TTCTAGCTAC	CATGGTGGAA	CAGCAAGTAG	17400
TGGTGTATC	AATATTAAAC	AAGATGTGAC	TCAAGATATC	AAAGGAAGAC	ATGTTCTATT	17460
TGTAGAAGAT	ATCATTGATA	CAGGTCAAAC	TTTGAAGAAT	TTGCCAGATA	TGTTTAAAGA	17520
AAGAGAAGCA	GCTTCTGTTA	AAATTGCAAC	CTTGTTGGAT	AAACCAGAAG	GACGTGTTGT	17580
AGAAATTGAG	GCAGACTATA	CTTGCTTTAC	TATCCCAAAT	GAGTTTGTAG	TAGGTTATGG	17640
TTTAGACTAC	AAAGAAAATT	ATCGTAATCT	TCCTTATATT	GGAGTATTGA	AAGAGGAAGT	17700
GTATTCAAAT	TAGAAAGAAT	AATCTTTAAT	GAAAAAACAA	AATAATGGTT	TAATTAATAA	17760
TCCTTTTCTA	TGGTTATTAT	TTATCTTTTT	CCTTGTGACA	GGATTCCAGT	ATTTCTATTC	17820
TGGGAATAAC	TCAGGAGGAA	GTCAGCAAAT	CAACTATACT	GAGTTGGTAC	AAGAAATTAC	17880
CGATGGTAAT	GTAAGAAGAT	TAACCTACCA	ACCAAATGGT	AGTGTTATCG	AAGTTTCTGG	17940
TGTCTATAAA	AATCCTAAAA	CAAGTAAAGA	AGAAACAGGT	ATTGAGTTTT	TCACGCCATC	18000
TGTTACTAAG	GTAGAGAAAT	TTACCAGCAC	TATTCTTCCT	GCAGATACTA	CCGTATCAGA	18060
ATTGCAAAAA	CTTGCTACTG	ACCATAAAGC	AGAAGTAACT	GTTAAGCATG	AAAGTTCAAG	18120

268

TGGTATATGG ATTAATCTAC TCGTATCCAT TGTGCCATTT GGAATCTAT TCTTCTTCCT	18180
ATTCTCTATG ATGGGAAATA TGGGAGGAGG CAATGGCCGT AATCCAATGA GTTTTGGACG	18240
TAGTAAGGCT AAAGCAGCAA ATAAAGAAGA TATTAAAGTA AGATTTTCAG ATGTTGCTGG	18300
AGCTGAGGAA GAAAAACAAG AACTAGTTGA AGTTGTTGAG TTCTTAAAAG ATCCAAAACG	18360
ATTACAAAA CTTGGAGCCC GTATTCCAGC AGGTGTTCTT TTGGAGGGAC CTCGGGGAC	18420
AGGTAAAAC TTGCTTGCTA AGGCAGTCGC TGGAGAAGCA GGTGTCCAT TCTTTAGTAT	18480
CTCAGGTTCT GACTTTGTAG AAATGTTTGT CGGAGTTGGA GCTAGTCGTG TTCGCTCTCT	18540
TTTGTAGGAT GCCAAAAAAG CAGCACCAGC TATCATCTTT ATCGATGAAA TTGATGCTGT	18600
TGGACGTCAA CGTGGAGTCG GTCTCGGCGG AGGTAATGAC GAACGTGAAC AAACCTTGAA	18660
CCAACTTTTG ATTGAGATGG ATGGTTTGA GGGAAATGAA GGGATTATCG TCATCGCTGC	18720
GACAAACCGT TCAGATGTAC TTGACCCTGC CCTTTTGCCT CCAGGACGTT TTGATAGAAA	18780
AGTATTGGTT GCTCGTCTG ATGTTAAAGG TCGTGAAGCA ATCTTGAAAG TTCACGCTAA	18840
GAATAAGCCT TTAGCAGAAG ATGTTGATTT GAAATTAGTG GCTCAACAAA CTCCAGGCTT	18900
TGTTGGTGCT GATTAGAGA ATGCTTGAA TGAAGCAGCT TTAGTTGCTG CTCGTCGCAA	18960
TAAATCGATA ATTGATGCTT CAGATATTGA TGAAGCAGAA GATAGAGTTA TTGCTGGACC	19020
TTCTAAGAAA GATAAGACAG TTTCACAAAA AGAACGAGAA TTGGTTGCTT ACCATGAGGC	19080
AGGACATACC ATTGTTGGTC TAGTCTTGTC GAATGCTCGC GTTGTCCTATA AGGTTACAAT	19140
TGTACCACGC GGCCGTGCAG GCGGATACAT GATTGCACTT CCTAAAGAGG ATCAAATGCT	19200
TCTATCTAAA GAAGATATGA AAGAGCAATT GGCTGGCTTA ATGGGTGGAC GTGTAGCTGA	19260
AGAAATTATC TTTAATGTCC AAACCACAGG AGCTTCAAAC GACTTTGAAC AAGCGACACA	19320
AATGGCACGT GCAATGCTTA CAGAGTACGG TATGAGTGAA AAACCTGGCC CAGTACAATA	19380
TGAAGGAAAC CATGCTATGC TTGCTGCACA GAGTCCCTCA AAATCAATTT CAGAACAAAC	19440
AGCTTATGAA ATTGATGAAG AGGTTCTGTC ATTATTAAAT GAGGCACGAA ATAAAGCTGC	19500
TGAAATTATT CAGTCAAATC GTGAAACTCA CAAGTTAATT GCAGAAGCAT TATTGAAATA	19560
CGAAACATTG GATAGTACAC AAATTAAAGC TCTTTACGAA ACAGGAAAGA TGCCTGAAGC	19620
AGTAGAAGAG GAATCTCATG CACTATCCTA TGATGAAGTA AAGTCAAAAA TGAATGACGA	19680
AAAATAACCC TGAGAGAGGC TGGAGCCTCT CTTTTTTGTG CAGTTTAGGA GCTAAAGGGA	19740
ACAGAATGGA GAAATGGAA CAAATGTGTT TTCTAATCTG TTAGACTGTA TCTAGAAAGG	19800
GGAAATTAT GATTAAAGAA TTGTATGAAG AAGTCCAAGG GACTGTGTAT AAGTGATGAA	19860
ATGAATATTA CCTTCATTA TGGGAATTGT CGGATTGGGA GCAAGAAGGC ATGCTCTGCT	19920

269

TACATGAATT GATTAGTAGA GAAGAAGGAC TGGTAGACGA TATTCCACGT TTAAGGAAAT	19980
ATTTCAAGAC CAAGTTTCGA AATCGAATTT TAGACTATAT CCGTAAACAG GAAAGTCAGA	20040
AGCGTAGATA CGATAAAGAA CCCTATGAAG AAGTGGGTGA GATCAGTCAT CGTATAAGTG	20100
AGGGGGGTCT CTGGCTAGAT GATTATTATC TCTTTCATGA AACACTAAGA GATTATAGAA	20160
ACAAACAAAG TAAAGAGAAA CAAGAAGAAC TAGAACGCGT CTTAAGCAAT GAACGATTTC	20220
GAGGGCGTCA AAGACTATTA AGAGACTTAC GCATTCTGTT TAAGGAGTTT ACTATCCGTA	20280
CCCACTAGTA AGTCATGCAA AAAAAATGAA AAAAATTAGA AAAAGTAGTT GACAAAGTTT	20340
GAAAAGGCTG TATAATAGTA AGAGTTGAAA ATAACAACCTC AGGTCCGTTG GTCAAGGGGT	20400
TAAGACACCG CCTTTTCACG GCGGTAACAC GGGTTCGAAT CCCGTACGGA CTATGGTATG	20460
TTGCGTCAGG ACCACTTGAT GAAAAAAGT TAAAAAAAC TAAAAATCT TCAAAAAAGT	20520
GTTGACAAGC GAAAGCAGTT GTGATATACT AATATAGTTG TCGCTTGAGA GAAGCAAGTG	20580
ACAAAGACCT TTGAAACTG AACAAGACGA ACCAATGTGC AGGGCGCTAC AACGTAAGTT	20640
GTAGTACTGA ACAATGAAAA AAACAATAAA TCTGTCAGTG ACAGAAATGA GTAAGAACTC	20700
AAACTTTTTA ATGAGAGTTT GATCCTGGCT CAGGACGAAC GCTGGCGGCG TGCCTAATAC	20760
ATGCAAGTAG AACGCTGAAG GAGGAGCTTG CTTCTCTGGA TGAGTTGCGA ACGGGTGAGT	20820
AACGCGTAGG TAACCTGCCT GGTAGCGGGG GATAACTATT GGAAACGATA GCTAATACCG	20880
CATAAGAGTA GATGTTGCAT GACATTTGCT TAAAAGGTGC ACTTGCAATCA CTACCAGATG	20940
GACCTGCGTT GTATTAGCTA GTTGGTGGG TAACGGCTCA CCAAGGCGAC GATACATAGC	21000
CGACCTGAGA GGGTGATCGG CCACACTGGG ACTGAGACAC GGCCAGACT CCTACGGGAG	21060
GCAGCAGTAG GGAATCTTCG GCAATGGACG GAAGTCTGAC CGAGCAACGC CGCGTGAGTG	21120
AAGAAGGTTT TCGGATCGTA AAGCTCTGTT GTAAGAGAAG AACGACTGTC AGAGTGAAAA	21180
GTTCACTAGT TGACGGTATC TTACCAGAAA GGGACGGCTA ACTACGTGCC AGCAGCCGCG	21240
GTAATACGTA GGTCCCAGC GTTGTCGGA TTTATTGGGC GTAAAGCGAG CGCAGGCGGT	21300
TAGATAAGTC TGAAGTAAA GGCTGTGGCT TAACCATA	21338

(2) INFORMATION FOR SEQ ID NO: 21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6273 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGTTTTTAAA GAGCCGTGTC TGGATAGACT TTCGGACGCA ACGCTCTATT AGATAATGAA	60
CTGCCTATAC ACAAGATTTC TAACCTTAGT CGACATGAGC TGAAACCTCT TATTTGTTAA	120
GTAGTTCACA AAATATTATA CACCTATTTT ATGAATAGTC AACTGTCTTT ACAGTAAAT	180
TTTAGAAAAAT CATGAAAATT TTCTCTTTCT TTCCATTTTA AGTGACATTC AGTCATTCTC	240
ACATCAAAAA AGCCCAGACG AAATTGTCTG AGCATTCTTT TATCTAGTCG TTTAAGGAAG	300
TTGAGTTCAG TATGTTTAAA GTCTCTGTCC CATCATTTCT TCAACAAACC TTGTTCTTGG	360
AGAAACTCCT TGGCTACTTG CTTTGCTGAC TTGCCTTCAA CACCGACTTG GTAGTTGAGC	420
TGGCTCATCT GGCTTTCTGT AATCTTACCA GCCAATGTAT TAAGAACTCT TTCCAACCTCT	480
GGGTGTTTCT TGAGAAGAGC TTCTTTCATG AGTGGAGCCC CTTGATAAGG TGGGAAGAGT	540
TGCTTGTCAT CTTCCAAGAC CTGTAAATCA TAACGCTCCA ATTCCGCATC AGTCGAATAG	600
GCATCCGTGA TTTGAATATC CCCTGACTGA ATAGCCTGAT AGCGAAGGGC TGGCTCAATG	660
GTCGCTACAT TGAGATTGAG ACCATACATT GATTGCAAGC CCTTATTTCC ATCTTCACGG	720
TCGTAAACT CGAGTGTAAG ACCTGCCTTC AACTGCCCTT CCACTTTTTT CAAGTCTGAA	780
ATGGTCTTCA AGCCATATTC TTGAGCAATC TTTTTCGGAA CAGCTACAGC ATAGGTGTTT	840
TGATAAGACA TGGGTTTGAG ATAGGCTAGA TGATCCTGCT TAGCAATGCC ATCACGCCCC	900
ACCTGATAAA CCTGTTCTGG TTCATGACTC ACCTTGGGTG ATGGTTGAAG CAAACTTTCA	960
GTCACCGTAC CAGTAAATTC AGGATAGATG TCAATATCGC CTTTTTTCAG AGCTTCATAA	1020
AGGAAGCTTG TCTTCCCAA ATTCGGTTTA ACAGTCGCAG TCATGCTGGT ATTTTCTTCA	1080
ATCAGCAACT TATACATATT GGCCAAAATT TCTGGTTCTG GACCTATTTT CCCAGCAATA	1140
ACCAAGTTTT CTTCTCTTT TTGAACCAA AGAGCTGGAC TATAAGACAG ACCCAGTAAT	1200
AAAGCCACCA AGGCAAAACC TGAGAAAATC GTCCGTAATT TTGCTTTTTT CATCACTTTT	1260
AGTAGGAAAT TAAAGGCAAT GGCTAGCACT GCAGAAGAAA GTGCCCAAT CAAAATCAAA	1320
CTGGCATTAT TACGGTCAAT TCCCAAAAGA ATAAAGGAAC CTAGTCCCCC TGCACCAATC	1380
AAGGCCGCCA AGGTGCGGT ACCGATAATC AAAACAGCTG CCGTCCGAAT CCCAGACATG	1440
ATAACAGGCA TGGCGAGTGG AATTTCAAAT TTCTTGAGAC GTTCCCATCT GGTCAATCCA	1500
AAGGCAATCC CAGCCTCTTG CAGGTTCGGA TCAATTCCCT TCAGCCAGT GATAGTATTT	1560
TGCAAAATAG GGAAAATCGC ATAAATCACT AGAGCTGTCA AAGCCGGCAA GGTCCCAATT	1620
CCCATCAAAG GGATAAAGAG CCCCAACAAG GCCAGAGACC GGATGGTCTG GAAAATACCT	1680
GCAATCTGCA AGACCCAGTC GGCCAGCTTC TCATGATAGC GAAGAAAAAC AGCCAAGGGA	1740

271

ATCGCAAGCA AAATAGCTAG TAACAAGGTC AAAAGCGACA ACTGCAAATG TTGAGATAGA	1800
GCTGTCAACC AATCACTAAA ACGATCCTGA AAAGTTGCAA TTAAATTAGT CATGAACACT	1860
ACCTCCAAAC AAGTCTGCTA CAAAGTCTGT TGCAGGCGCT TTAAAAATTG TCTCGGGATT	1920
CGCTACCTGG CGAATTCTC CATCCTGCAA GACAGCAATA CCGTCCGCCA ACTTCAAGGC	1980
TTTCATCCGTA TCATGGGTTA CAAAAATCGT TGTTCATCCCA AACTCTTTAT GCAATTCCTT	2040
TGTCAGAACC TGCAACTGTT TTCTCGAAAT AGCATCCAAG GCCGAAAAGG GTTCATCCAT	2100
GAGGAAAATC TTGGGCTGAC CAATCATAGC TCGGACAATA CCGACCCGTT GCTGTTCTCC	2160
ACCAGATAAT TCACTAGGTA AGCGATGCCC ATACTCGGCT ACTGGTAAAC CAACCTTAGC	2220
CAAAAGCTCT TCTGTTTCT TCGTAATTTC TTCTTGCTC CACCCCTTCA TTTCAGGAAT	2280
GAGAGCAATA TTTTCCGCAA CTGTTAGATT TGGAAAAAGA GCAATAGCCT GTAAACATA	2340
ACCAGTAGAA AGACGAAGTT CACGCTCATC ATAGTCTTTG ATGCGCTTCC CATCCATATA	2400
AATATTTCCT TCAGTTGGTT CCAAAAGACG GTTAATCATC TTGAGCATGG TCGTCTTACC	2460
TGACCCAGAA GGCCCTACTA AAACCATAAA TTCCCATCC TCAATCTGTA AGTTGACATC	2520
TCTCAAGACA TCCTTTTCTG TGTAGCGCAG TGCTACATTT TTGTATTCAA TCATTCTTTG	2580
TCCTCAATTT AAAACTTCCC TCGATTGGTC AAGTCTTCTA CCTTAGGCAT AACTTCCTTA	2640
TTATCCCAAT GCTCCACAAT TTTCCCGTTC TCTAAACGGA AGATATCGTA CTGGGCATAA	2700
GCAACGCCAT CAATCTGAGT CTGACCATAG CTAACCACAT AGTTTCCTTG TCCTAAGAGT	2760
TGGAAAACAA AGTCAAAAGT GACACTATAT TCAGCCACAT AGTTTTTATA AGCAGCACTT	2820
CCTTGTCCTA TATCATGATT ATGCTGAATC AAATCGTCTG CCACATAATC ACTCCACTGC	2880
TCTAGCTCCC CATTTTGGAA AATTTCTCTC AAGAAACGGC GACCCAGCTT TCTATTTCTT	2940
GCTTTCTTAT CCAAACTCCTT GATTTCAAAA TCTCCAAAAA TTGATCTAG TTGGTCATTT	3000
TCAGGTGTTT GATAGTAGTC AATGACATCC CAATGCTCAA CAATACAACC ATTCTCATCC	3060
TCACGGAAAG TATCCGTCGT CACCCATTGA GCTTCTCCAC CATTGAGATA TTGATGAACA	3120
TGAACAAAGA CCAGATTGCC ATCCTCAATG GTGCGGACAA TCTTAATCTG ACGCTCTGGA	3180
TGACGCTCAA AGAAATCTGC AAAGAAGGCT GCAAATCCTT CTTTCCCGTC AGGAACACCT	3240
GTCGAATGTT GGATATAGGT ATCCCTTACA GACTGGGCTT GAGCCTCAGC AACTCGTCCG	3300
TCTTGAATGG CATGGATGTA TAGGTTGTGA GCATTTTCA CTTGTTGTGA CATATTCTAA	3360
ACCTCATTTT CTTCTCTTT CAGATTCGCC AAAATTCTTT CTTGAAAACC TTCAAAATTG	3420
TGAATTTCTT CCTCTGAAAA TCCTTTGTAA AAGATAGTAT CCAATTTCTG ACTGACACGA	3480

272

TGCCCCACTT CTTTCTGGGA CTTGCCTAAC TCCGTTAAAA CTAAATACTT CTTACGCTTG	3540
TCTTTTCCAC ACGGACTAAC AATTACAAGC TTTTGTTCCCT CTAGCTTTTT TATCATAGTC	3600
GTCAGCGTAT TATTGCAAG TCCAGTCGCA AGCGCGATAT CTGTCGCAGT TGGCAGCCA	3660
GTTTCACTAT TCCATAAAC CGCTAAAATC TTGCCCTGTT CACCCCTATA AAGAGCCTCA	3720
GGATCTTGAC TCAGTAACTT TTGAAAAATC CGCCCATTC ACAAACGAAT ATGATGGGCT	3780
AGCAAATGAC CATCTTTCAT AACACCTCCA ATTTATTTTCG ATATCGAAAT GAATAAAACA	3840
ATTGTAACAC TCATCGTTCT AACTGTCAAC TATTTTCGATT TAGAAATAAT TTTTGATAAT	3900
TATCCACACC ACCATACTCC GGCTCAACTA ACTTTTAACG AGAGTTTCTA AACTCCTTCG	3960
TCCTCCAGTC TACAAAAGCC TTCCATTCTG ACTATCCTAT ATTTTATGAG GGGACACATT	4020
TTTCCTATCA GACCATTTAT TTTAAAGATA GAAGTAAATC ATAATTGCTT CCATCTGTTT	4080
TTTTATAGTA TATTGAAGTT AGACTAGAGC ACTGTATCTT CTAAAACATT GATAGAAAGC	4140
GATTTGAATT TCCCAATCAA TTTGTTCTGA TTTATAGCAT TTCGAACTG GAATAGGACA	4200
CCATGACTGC TAAAAGATTT CTATAAATTC ATTTAATTTT CTCAATCAAT TTGTTTCATAT	4260
CTTATTTTCAT TCCGCTATAA TTTCACCTTA CCCTATCTTT TTCGTAGCAC CCTTCAAACA	4320
GCCTATCCCC TACCGTTTGA CGATTCTCTA CTTGCTCCA CTTCCATTAC AGAAGTTTCT	4380
TCACTACTAT GGGCTCGGCT GACTTCTCAT GATTCCTTGT TACTACTATT TGAACGCTCA	4440
CGAGATAGAT CTTACAAAAA ATGCTTTGAT CCACAATGGA ATCAAAGCAT TTTAAAGAGT	4500
TCCTCATACA TAAGCGCAGA AGTCGCAGTT CCTCTGTACT TGGCTTCTTC TCTTTTGACA	4560
AAGCGAGCCA AGTTGAGCAA CTCAGGTGCT GGATGTTTGG GATTTAGGAG CAATTCACGA	4620
TTGACCAGGC CTGAGAGACG AACTGCTCTG AATTGCTCAT TTGTAGTAGG CAGTTTTTTA	4680
GTAGTCTCTA GGAGAGCAGC AACTAAATCT TCACTCAAAT CATGTCGAGC ATGATTGTAA	4740
AGATCTTTTA TAAGGCTTTC TAGGTTTGGT TCTACCATCC CTACCACCTC CCTTATGGTT	4800
TAATAATGTT TAATCAAATC AACCGTTGAA CGATCCAATT TCTTCACCAA GGCTTGTAAG	4860
AAAGCTTGGC CTTCTAGGAA GTCATCCATT GCATAGAGGG TTTGGTGAGA ATGGATATAA	4920
CGAGCGCAGA CACCGATAGT TGTGATGGG ACACCACCAT TTTTCAGATG AGCTGCACCT	4980
GCATCTGTTT CGCCTTTACC ACAGTAGTAT TGGTACTTGA TACCAGCTTC TTCAGCCGTT	5040
GTCAAAGGA AATCCTTCAT CCCTGGGAGA AGCAAGTGAC CTGGATCATA GAAACGAATC	5100
AAGGTTCCAT CTCCAATCTT GCCTTGACCA CCGTAGACAT CACCTGCTGG TGAGCAATCA	5160
ACTGCGAGGA AGACTTCTGG GTCAAACCTG GTTGTAGAGG TATGAGCGCC ACGCAGACCA	5220
ACTTCTCTCT GGACGTTAGA ACCCAGATAG AGTTCATTGC CGAGTTTTTG ACCCGATAAA	5280

273

GCTTCAGCTA GCTCGCTTAC CATGAGGACA CCGTAGCGGT TATCCCAAGC TTTTGAGATG	5340
ATATTTTTTT CATTTGGCTGT CAAAATTGCA GAACTATCTG GTACAATGGT ATCACCAGGA	5400
CGGATGCCAA AACTTTCTGC CTCAGCCTTG TCCGCAAAAC CACCATCAAA AACGATATCG	5460
GCAATGGCTG GCATGGTTGG TCCCCCTTT CCACGAGTCA AATGCGGAGG AACAGAACCT	5520
GAAATCACAG GAATTTTCATG ACCATCACGA GTCAAGAGTT TGAAACGTTG GCTGCTAACC	5580
ACCATGGGGT TCCAGCCACC GATTTCTACG ACACGGAAGG TACCATCTGG CTTGATTTTCG	5640
CTGACCATAA AACCAACTTC GTCCATATGA GAAGCGACCA AGACGCGCGG TGCATCCACA	5700
GCTTCTGAAT GTTTGATACC AAAAATACCA CCCAAGCCAT CTGTCACCAC TTCATCCACA	5760
TGCGGTGTCA ACTTTTCACG AAGATAAGCA CGGACAGGCG CTTTCATGACC TGAGACTGCA	5820
GCAAGTTCTG TTAATTCTTT AATTTTTGAA AATAATGTTG TCATTTCACT TCCTTCTTTC	5880
TTTCATCCAT TTTACCACTT TTTATAGGAG AAGGATAGTG GGAAGGTGGA TTTCTAAGTT	5940
AGTATCTTAG TCCTGCTCTA TCTTAGAAAA GGATAGTATT CTCTTGCATG TAGTGCAAAA	6000
TCTAGTAAAC ATTCCAAAAT TAACTCGAAT ATTTATTTCC AAACAAAAAA ACAATACACC	6060
ATCAAAGTTG TTTGGATTTT TCATGAAATT TACAGAAAAT AGTTGACTTC CCTTTCTTCT	6120
TTCTTTAAAT ATATAGTTGG TTGAGTTTGG AATAGTACGC TGTAGCTGCT AAAACATTTT	6180
TAGAAATTAA TTTGACTTTC CTAATAGAGT TGTTCATATC TTATTTCAAT TTAATATAGT	6240
ACAAAACCTAG AAAAGGAAAA AATCATGACC AGG	6273

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28171 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

ACAACCTTTT TCAAAAACCTC ACCTTGGTAC GGAGATGTTT TGCTTTCTGC TATTATTTTC	60
GGTTATATTC ATATCAATTT TGCTTTAACT CCTCTTGCTT TTTTCATTTA TGCTAGTGGA	120
GGTCTTATTT TAGCTCTATT GTATCGCATG ACTAAAAATC TCTACTATCC AATACTAGTT	180
CATATTCTCA TTAATATCAC TGCCTTCTGG GATGTGTGGT TGCTCCTATT TTCAGGAAGT	240
TAGCTTACTA AAATAATGTC GGAACTTTCC GGCATTTTCT TTTTTCACAA ATAGTCAACG	300
TTTTTCTTTT CGATATTGTA GTGGTGTGTA TCCAGTTATT TTTTGAATT GATTTTGAAA	360



274

ATAAGGTTGA CTTGAGAAAG GCAGATAGTG AAGATAGTTA AGAAGAATAG GATGTTCTTT	420
TTTCCTTTTT GGAAAACTTC TAAAATATGG TATAATGAAA AGATAAAGAA GTTGGGGGTA	480
GAAGATGAAC ATTCAACAAT TACGCTATGT TGTGGCTATT GCCAATAGTG GTACTTTTTCG	540
TGAAGCTGCT GAAAAGATGT ATGTTAGTCA GCCGAGTCTG TCTATTTCTG TTCGTGATTT	600
GGAAAAAGAG TTGGGCTTTA AGATTTTCCG TCGGACCAGC TCAGGGACTT TCTTGACCCG	660
TCGTGGGATG GAATTTTATG AAAAATCGCA AGAATTGGTT AAAGGATTTG ATATTTTTC	720
AAATCAGTAT GCCAATCCTG AAGAAGAAAA AGATGAATTT TCTGTTGCTA GCCAGCACTA	780
TGACTTCTTG CCACCAACTA TTACGGCCTT TTCAGAGCGC TATCCTGACT ATAAGAACTT	840
CCGTATTTTT GAATCAACTA CTGTTCAAAT ATTAGATGAA GTGGCGCAAG GGCATAGTGA	900
GATTGGGATT ATCTACCTCA ACAATCAAAA TAAAAGGGG ATTATGCAAC GGGTTGAAAA	960
ATTAGGTCTG GAGGTATCG AATTGATTCC TTTCCATACC CATATTTATC TCCGTGAGGG	1020
TCATCCTTTA GCCCAGAAAG AGGAATTAGT CATGGAGGAT TTAGCGGATT TACCAACGGT	1080
TCGTTTCAC? CAAGAGAAAG ACGAGTACCT TTATTATTCA GAGAACTTTG TCGATACCAG	1140
CGCTAGCTCA CAGATGTTTA ATGTGACAGA CCGTGCCACC TTGAATGGTA TTTTGGAGCG	1200
GACGGACGCC TATGCGACAG GTTCTGGATT TTTAGATAGT GACAGTGTTA ATGGCATTAC	1260
AGTTATTCGT CTCAAGGATA ACCTAGATAA CCGCATGGTC TATGTTAAAC GTGAAGAAGT	1320
GGAGCTTAGT CAAGCTGGGA CTCTCTTCGT AGAAGTCATG CAAGAATATT TTGATCAAAA	1380
GAGGAAATCA TGAAAAAAG AGCAATAGTG GCAGTCATTG TACTGCTTTT GATTGGGCTG	1440
GATCAGTTGG TCAAA?CCTA TATCGTCCAG CAGATTCCAC TGGGTGAAGT GCGCTCCTGG	1500
ATCCCCAATT TCGTTAGCTT GACCTACCTG CAAAATCGAG GTCCAGCCTT TTCTATCTTA	1560
CAAGATCAGC AGCTGTTATT CGCTGTCATT ACTCTGGTTG TCGTGATAGG TGCCATTTGG	1620
TATTTACATA AACACATGGA GGACTCATTC TGGATGGTCT TGGGTTTGAC TCTAATAATC	1680
GCGGGTGGTC TTGGAACTT TATTGACAGG GTCAGTCAGG GCTTTGTTGT GGATATGTTT	1740
CACCTTGACT TTATCAACTT TGCAATTTTC AATGTGGCAG ATAGCTATCT GACGGTTGGA	1800
GTGATTATTT TATTGATTGC AATGCTAAAA GAGGAAATAA ATGGAAATTA AAATTGAAAC	1860
TGGTGGTCTG CGTTTGGATA AGGCTTTGTC AGATTTGTC	1920
GAATGAACAA ATTAAATCAG GCCAGGTCTT GGTCATGGT CAAGTCAAGA AAGCTAAATA	1980
CACAGTCCAA GAGGGTGATG TCGTCACTTA CCATGTGCCA GAACCAGAGG TATTAGAGTA	2040
TGTGGCTGAG GATCTTCCGC TAGAAATAGT CTACCAAGAT GAGGATGTGG CTGTCGTTAA	2100
CAAACCTCAG GGAATGGTTG TGCACCCGAG TGCTGGTCAT ACCAGTGGA	2160

275

TGCCCTCATG TATCATATTA AGGACTTGTC GGGTATCAAT GGGGTTCTGC GTCCAGGGAT	2220
TGTTACCCGT ATTGATAAGG ATACGTCAGG TCTTCTCATG ATTGCTAAAA ACGATGATGC	2280
GCATCTAGCA CTTGCCCAAG AACTCAAGGA TAAAAAGTCT CTCCGCAAA ATTGGGCGAT	2340
TGTTTCATGGA AATCTACCTA ATGATCGTGG TGTAAATTGAA GCGCCGATTG GCCGGAGTGA	2400
AAAAGACCGT AAGAAACAGG CTGTAAGTGC TAAAGGGAAG CCTGCAGTGA CGCGTTTTC	2460
CGTCTTGGA CGCTTTGGCG ATTATAGCTT AGTAGAGTTG CAACTGGAGA CAGGGCGCAC	2520
TCATCAAATC CGTGTCCACA TGGCTTATAT CGGCCATCCA GTCGCTGGTG ATGAGGTCTA	2580
TGGTCCTCGC AAGACTTTGA AAGGACATGG ACAATTTCTT CATGCCAAGA CTTTAGGTTT	2640
TACTCATCCG AGAACAGGTA AGACCTTGA ATTTAAAGCA GATATCCCAG AGATTTTAA	2700
GGAAACCTTG GAGAGATTGA GAAAGTAAGA ATGAAAAAGA AATTAAGTAG TTTAGCACTT	2760
GTAGGCGCTT TTTTAGGTTT GTCATGGTAT GGAATGTTT AGGCTCAAGA AAGTTCAGGA	2820
AATAAAAATCC ACTTTATCAA TGTTCAGAA GGTGGCAGTG ATGCGATTAT TCTTGAAAGC	2880
AATGGACATT TTGCCATGGT GGATACAGGA GAAGATTATG ATTTCCAGGA TGGAAGTGAT	2940
TCTCGCTATC CATGGAGAGA AGGAATTGAA ACGTCTTATA AGCATGTTCT AACAGACCGT	3000
GTCTTTCGTC GTTTGAAGGA ATTGGGTGTC CAAAAACTTG ATTTTATTTT GGTGACCCAT	3060
ACCCACAGTG ATCATATTGG AAATGTTGAT GAATTACTGT CTACCTATCC AGTTGACCGA	3120
GTCTATCTTA AGAAATATAG TGATAGTCGT ATTACTAATT CTGAACGCT ATGGGATAAT	3180
CTGTATGGCT ATGATAAGGT TTTACAGACT GCTGCAGAAA AAGGTGTTTC AGTTATTCAA	3240
AATATCACAC AAGGGGATGC TCATTTTCAG TTTGGGGACA TGGATATTCA GCTCTATAAT	3300
TATGAAAATG AACTGATTC ATCGGCTGAA TTAAGAAAA TTTGGGATGA CAATTCCAAT	3360
TCCTTGATTA GCGTGGTGAA AGTCAATGGC AAGAAAATTT ACCTTGGGGG CGATTTAGAT	3420
AATGTTTCATG GAGCAGAAGA CAAGTATGGT CCTCTCATTG GAAAAGTTGA TTTGATGAAG	3480
TTTAATCATC ACCATGATAC CAACAAATCA AATACCAAGG ATTTCAATTA AAATTTGAGT	3540
CCGAGTTTGA TTGTTCAAAC TTCGGATAGT CTACCTTGGA AAAATGGTGT TGATAGTGAG	3600
TATGTTAATT GGCTCAAAGA ACGAGGAATT GAGAGAATCA ACGCAGCCAG CAAAGACTAT	3660
GATGCAACAG TTTTGTATAT TCGAAAAGAC GGTTTTGTCA ATATTCAAC ATCCTACAAG	3720
CCGATTCCAA GTTTTCAAGC TGTTTGGCAT AAGAGTGCAT ATGGGAACTG GTGGTATCAA	3780
GCGCCTGATT CTACAGGAGA GTATGCTGTC GGTGGAATG AAATCGAAGG TGAATGGTAT	3840
TACTTTAACC AAACGGGTAT CTGTGTACAG AATCAATGGA AAAAATGGAA CAATCATTTG	3900

276

TTCTATTTGA CAGACTCTGG TGCTTCTGCT AAAAATTGGA AGAAAATCGC TGGAAATCTGG	1960
TATTATTTTA ACAAAGAAAA CCAGATGGAA ATTGGTTGGA TTCAAGATAA AGAGCAGTGG	4020
TATTATTTGG ATGTTGATGG TTCTATGAAG ACAGGATGGC TTCAATATAT GGGGCAATGG	4080
TATTACTTTG CTCCATCAGG GGAAATGAAA ATGGGCTGGG TAAAAGATAA AGAAACCTGG	4140
TACTATATGG ATTCTACTGG TGTCATGAAG ACAGGTGAGA TAGAAGTTGC TGGTCAACAT	4200
TATTATCTGG AAGATTCAGG AGCTATGAAG CAAGGCTGGC ATAAAAAGGC AAATGATTGG	4260
TATTTCTACA AGACAGACGG TTCACGAGCT GTGGGTTGGA TCAAGGACAA GGATAAATGG	4320
TACTTCTTGA AAGAAAATGG TCAATTACTT GTGAACGGTA AGACACCAGA AGGTTATACT	4380
GTGGATTCAA GTGGTGCCTG GTTACTGGAT GTTTCGATCG AGAAATCTGC TACAATTAAA	4440
ACTACAAGTC ATTCAGAAAT AAAAGAATCC AAAGAAGTAG TGAAAAAGGA TCTTGAAAAAT	4500
AAAGAAACGA GTCAACATGA AAGTGTTACA AATTTTTCAA CTAGTCAAGA TTTGACATCC	4560
TCAACTTCAC AAAGCTCTGA AACGAGTGTA AACAAATCGG AATCAGAACA GTAGTAGAAA	4620
AGAAGGTTTT AGGGCCTTCT TTTTCCTATC AACTCTTTTC TATTTCTGT TATTCATGTT	4680
ATAATGGATA AATATGAATA ATCGGAGTGA GACTATGAAA TACAAACCGA TTGTCTTTAA	4740
GGTGGTACT TCTTCTCTGA CAAATGAGGA TGGAAAGTTA TCACGTAGTA AGGTAAAGGA	4800
TATTACCCAG CAGTTGGCTA TGCTGCACGA GGCTGTCAT GAGTTGATTT TGGTGTCTTC	4860
AGGTGCCATT GCGGCTGGTT TTGGAGCCTT AGGATTTAAA AAGCGTCCGA CTAAGATTCC	4920
TGATAAACAG GCTTCAGCAG CCGTAGGGCA GGGGCTTTTG TTGGAAGAAT ATACAACCAA	4980
TCTTCTCTG CGTCAAAATCG TTTCTGCACA AATCTTGCTG ACCCAAGATG ACTTTGTGGA	5040
TAAGCGTCGT TATAAAAATG CCCATCAGGC TTGTGCGGTT TTGCTCAACC GTGGGGCAAT	5100
TCCTATCATC AATGAGAATG ATAGTGTCTG TATTGATGAG CTCAGGTTG GGGACAATGA	5160
CACCTAAGT GCTCAAGTAG CGCGATGGT CCAAGCAGAC CTTTGTAGTT TCTTGACAGA	5220
TGTCGACGGT CTCTATACTG GAAATCCTAA TTCAGATCCA AGAGCCAAAC GCTTGGAGAG	5280
AATCGAGACC ATCAATCGTG AGATTATTGA TATGGCTGGT GGAGCTGGT CGTCAAACGG	5340
AACTGGGGT ATGTTAACCA AAATCAAGGC TGCAACTATC GCGACGGAAT CAGGAGTTCC	5400
TGTTTATATC TGCTCATCCT TGAATCAGA TTCCATGATT GAGGCGGCAG AGGAGACCGA	5460
GGATGGTTCT TACTTTGTG CTCAAGAGAA GGGGCTTCGT ACCCAGAAAC AATGGCTTGC	5520
CTTCTATGCT CAGAGTCAAG GTTCTATTTG GGTGATAAAA GGGGCTGCGG AAGCTCTCTC	5580
TCAATATGGA AAGAGTCTTC TCTTATCTGG TATCGTTGAA GCAGAAGGAG TCTTTCTTA	5640
CGGTGATATC GTGACAGTAT TTGACAAGGA AAGTGGAAAA TCACTTGGAA AAGGACGCGT	5700

277

GCAATTTGGA	GCATCTGCTT	TGGAGGATAT	GTTGCGTTCT	CAAAAAGCCA	AGGGTGTCTT	5760
GATTTACCGT	GACGACTGGA	TTTCCATTAC	TCCTGAAATC	CAACTACTTT	TTACAGAATT	5820
TTAGAGGTAA	ACTATGGTGA	GTAGACAAGA	ACAAATTTGAA	CAGGTACAGG	CTGTTAAAAA	5880
ATCGATTAAC	ACAGCTAGTG	AAGAAGTGAA	AAACCAAGCC	TTGCTAGCCA	TGGCTGATCA	5940
CTTAGTGGCT	GCTACTGAGG	AAATTTTAGC	GGCTAATGCC	CTCGATATGG	CAGCGGCTAA	6000
GGGGAAAATC	TCAGATGTGA	TGTTGGATCG	TCTTTATTTG	GATGCAGATC	GTATAGAAGC	6060
GATGGCAAGA	GGAATTCGTG	AAGTGGTTGC	CTTACCAGAT	CCAATCGGTG	AAGTTTTAGA	6120
AACAAGTCAG	CTTGAAAATG	GTTTGGTTAT	CACAAAAAAA	CGTGTAGCTA	TGGGTGTCAT	6180
CGGTATTATC	TATGAAAGCC	GTCCAAATGT	GACGCTCGAT	GCGGCTGCTT	TGACTCTTAA	6240
GAGTGGAAAT	GCGGTTGTTT	TTCGTAGTGG	TAAGGATGCC	TATCAAACAA	CCCATGCCAT	6300
TGTCACAGCC	TTGAAGAAGG	GCTTGGAGAC	GACTACTATT	CATCCAAATG	TGATTCAACT	6360
GGTGGAGGAT	ACTAGCCGTG	AAAGTAGTTA	TGCTATGATG	AAGGCCAAGG	GCTATCTAGA	6420
CCTTCTCATT	CCTCGTGGAG	GAGCTGGCTT	GATCAATGCA	GTGGTTGAGA	ATGCGATTGT	6480
ACCTGTTATC	GAGACAGGGA	CTGGGATTGT	CCATGCTCTAT	GTGGATAAGG	ATGCAGACGA	6540
AGACAAGGCG	CTGCTCTATCA	TCAACAATGC	TAAAACCACT	CGTCCTTCTG	TTTGTAATGC	6600
CATGGAGGTT	CTGCTGGTTC	ATGAAAACAA	GGCAGCAAGC	TTCTTCCTC	GCTTGGAGCA	6660
AGTGTGGTTT	GCAGAGCGTA	AGGAAGCTGG	ACTGGAACCA	ATTCAATTCC	GCCTAGATAG	6720
CAAAGCAAGC	CAGTTTGTTT	CAGGTCAAGC	AGCTGAGACC	CAAGACTTTG	ACACCGAGTT	6780
TTTAGACTAT	GTCCTTGCTG	TTAAGGTTGT	GAGCAGTTTA	GAAGAAGCGG	TTGCCGACAT	6840
TGAATCCCAC	AGCACCCATC	ATTCCGATGC	TATTGTGACG	GAAAATGCTG	AAGCTGCAGC	6900
ATACTTTACA	GATCAAGTGG	ACTCTGCAGC	GGTGTATGTT	AATGCCTCAA	CTCGTTTCAC	6960
AGATGGAGGA	CAATTTGGTC	TTGGTTGTGA	AATGGGGATT	TCTACTCAGA	AATTGCACCC	7020
GCGTGGTCCC	ATGGGCTTGA	AAGAGTTGAC	CAGCTACAAG	TATGTGGTTG	CCGGTGATGG	7080
GCAGATAAGG	GAGTAAGAGA	TGAAGATTGG	ATTTATCGGT	TTGGGGAATA	TGGGTGCTAG	7140
CTTGGCCAAAA	TCTGTCTTGC	AGACTAGGAC	GTCAGATGAG	ATTCTCCTTG	CCAATCGTAG	7200
TCAAGCTAAG	GTAGATGCTT	TCATTGCAGA	CTTTGGTGGT	CAGGCTTCCA	GCAATGAAGA	7260
AATGTTTGCA	GAAGCAGATG	TGATTTTCT	AGGAGTTAAG	CCTGCTCACT	TTTCTGAACT	7320
GCTTTCTCAA	TACCAGACCA	TCCTTGAAAA	AAGAGAAAGT	CTTCTTTTGA	TTTCGATGGC	7380
AGCTGGATTG	ACCTTAGAAA	AACTAGCAAG	TCTTATCCCA	AGTCAACACC	GAATTATTGG	7440

278

TATGATGCCT AATACCCCTG CTTCTATCGG GCAAGGAGTG ATTAGTTATG CCTTGTCTCC	7500
TAATTGCAGG GCTGAGGACA GTGAGCTCTT TTATCAGCTT TTAGCCAAGG CTGGTCTCTT	7560
GGTTGAACTA GGAGAAAGTT TAATCGATGC AGCGACAGGT CTTGCAGGTT GTGGACCAGC	7620
CTTTGTCTAT CTTTATATCG AGGCCTTGGC AGATGCAGGT GTTCAGACAG GATTACCACG	7680
AGAAATAGCA TTGAAAATGG CAGCACAAAC TGTGGTAGGA GCTGGGCAAT TGCTCCTTGA	7740
AAGTCAGCAA CATCCTGGAG TATTGAAAGA CCAAGTCTGT AGCCCAGGCG GTTCGACTAT	7800
CGCTGGTGTA GCAAGCCTAG AAGCGCATGC TTTCCGAGGA ACAGTCATGG ATGCAGTTCA	7860
TCAAGCCTAC AAACGAACAC AAGAACTAGG TAAATAAGAG GTAGTTTGA CTGCCCTCTT	7920
TATGGTGGCT GAAATGAGAA GACACAAAAA GATTGTCACA AACCCCTATT TTTTGTATAG	7980
AATAGAAGTA GTAAAAAGA AATGAGTTAG ACATGTCAAA AGGATTTTGA GTCTCTCTTG	8040
AGGGACCAGA GGGAGCAGGC AAGACCAGTG TTTTAGAGGC TCTGCTACCA ATTTTAGAGG	8100
AAAAAGGAGT AGAGGTGTTG ACGACCCGTG AACCTGGCGG AGTCTTGATT GGGGAGAAGA	8160
TTCCGGGAAGT GATTTTGGAT CCAAGTCATA CTCAGATGGA TGCTAAAACA GAGCTACTTC	8220
TCTATATTGC CAGTCGCAGA CAGCATTTGG TGGAAAAAGT TCTTCCAGCC CTTGAAGCTG	8280
GCAAGTGGT CATCATGGAT CGTTTATCG ATAGTTCTGT TGCCTATCAG GGATTTGGTC	8340
GTGGCTTAGA TATTGAAGCC ATTGACTGGC TCAATCAGTT TCGCAGAGAT GGCCTCAAAC	8400
CCGATTGAC ACTCTATTTT GACATCGAGG TGAAGAAGG GCTGGCTCGT ATTGCTGCTA	8460
ATAGTGACCG CGAGGTAAAT CGTTTGGATT TGAAGGGTT GGAATTGCAT AAAAAAGTTC	8520
GTCAAGGCTA CTTTCTCTT CTGGATAAAG AGGGAAATCG CATTGTCAAG ATTGATGCTA	8580
GTCTCCCTTT GGAGCAAGTT GTGGAACTA CCAAGGCTGT CTGTTTGAC GGAATGGGCT	8640
TGGCCAAATG AAACAAGATC AACTAAAGGC TTGGCAACCA GCTCAGTTTG ACCGTTTTGT	8700
CCGTATCTTA GAACAAGACC AGCTCAATCA CGCCTATCTC TTTTCAGGTT TCTTTGAAAG	8760
CTTGGAATG GCGCAATTTT TAGCTAAGAG CCTCTTTTGT ACGGATAAAG TTGGCGTCTT	8820
ACCATGTGAG AAATGCCGAA GTTGCAAGCT GATTGAACAG GGAGAATTTT CCGATGTCAC	8880
CTTGATTAAA CCAGTTAATC AGGTCAATTA GACGGAACGC ATTCGAGAAT TGGTGGGTCA	8940
GTTTTCTCAA GCAGGGATTG AAAGCCAGCA ACAGTCTTTT ATCATCGAGC AAGCGGATAA	9000
AATGCATCCC AACGCAGCCA ATTCTCTGCT CAAGGTCATC GAAGAACCCC AGAGTGAAGT	9060
TTATATTTTC TTCTTGACTA GCGATGAGGA AAAGATGTTA CCGACAATCC GAAGTCGGAC	9120
TCAGATCTTC CACTTTAAAA AGCAAGAAGA AAACTTATC TTA CTCTTAG AACAAATGGG	9180
ACTTGTTAAG AAAAAAGCCA CTCTTTTAGC TAAGTTTAGT CAATCGCGAG CTGAAGCAGA	9240

AAAGTTGGCT AATCAGGCAA GTTTTGGAC CTTGGTCGAT GAAAGTGAAC GCCTGCTGAC	9300
TTGTTAGTA GCTAAGAAAA AAGAAAGTTA TCTACAGGTT GCCAAATTAG CCAACTTGGC	9360
AGATGATAAG GAAAAACAGG ATCAGGTTTT ACGGATTCTT GAAGTTCTCT GTGGGCAGGA	9420
CCTCTTGCAG GTAAGAGTAA GAGTGATTCT ACAAGATTTA CTAGAAGCTA GAAAAATCTG	9480
GCAAGCTAAT GTCAGCTTTC AAAATGCCAT GGAATATCTG GTCTTGAAAG AAATATAAAC	9540
TCAAAAATGA ATGATAAAGA AAGGAAAGGG CTGTTTTATG GACAAAAAAG AATTATTTGA	9600
CGCGCTGGAT GATTTTCCC AACAATTATT GGTAACCTTA GCCGATGTGG AAGCCATCAA	9660
GAAAAATCTC AAGAGCCTGG TAGAGGAAAA TACAGCTCTT CGCTTGGAAA ATAGTAAGTT	9720
GCGAGAACGC TTGGGTGAGG TGGAAGCAGA TGCTCCTGTC AAGGCCAAGC ATGTTCTGTA	9780
AAGTGTCGT CGCATTTACC GTGATGGATT TCACGTATGT AATGATTTTT ATGGACAACG	9840
TCGAGAGCAG GACGAGGAAT GTATGTTTTG TGACCAGTTG CTATACAGGG AGTAGGCATG	9900
CAGATTCAAA AAAGTTTTAA GGGGCAGTCT CCCTATGGCA AGCTGTATCT AGTGGCAACG	9960
CCGATTGGCA ATCTAGATGA TATGACTTTT CGTGCTATCC AGACCTTGAA AGAAGTGGAC	10020
TGGATTGCTG CTGAGGATAC GCGCAATACA GGGCTTTTGC TCAAGCATTT TGACATTTCC	10080
ACCAAGCAGA TCAGTTTCA TGAGCACAAAT GCCAAGGAAA AAATTCCTGA TTTGATTGGT	10140
TTCTTGAAAG CAGGGCAAAG TATTGCTCAG GTCTCTGATG CCGGTTTGCC TAGCATTTCA	10200
GACCCTGGTC ATGATTTAGT TAAGGCAGCT ATTGAGGAAG AAATTGCAGT TGTGACAGTT	10260
CCAGGTGCCT CTGCAGGAAT TTCTGCCTTG ATTGCCAGTG GTTTAGCGCC ACAGCCACAT	10320
ATCTTTTACG GTTTTTTACC GAGAAAATCA GGTACAGAGA AGCAATTTTT TGGCTTGAAA	10380
AAAGATTATC CTGAAACACA GATTTTTTAT GAATGAGCTC ATTGTGTAGC AGACACGTTG	10440
GAAAAATATGT TAGAAGTCTA CGGTGACCGC TCCGTTGTCT TGGTCAGGGA ATTGACCAAA	10500
ATCTATGAAG AATACCAACG AGGTACTATC TCTGAGTTAT TAGAAAGCAT TGCTGAAACG	10560
CCACTCAAGG GCGAATGTCT TCTCATGTT GAGGGTGCCA GTCAGGGTGT GGAGGAAAAG	10620
GACGAGGAAG ACTTGTTCTG AGAAATTCAA ACCCGCATCC AGCAAGGTGT GAAGAAAAAC	10680
CAAGCTATCA AGGAAGTCGC TAAGATTTAC CAGTGAATA AAAGTCAGCT CTACGCTGCC	10740
TACCACGACT GGAAGAAAA ACAATAAAGG GAGACAGGAT GTAATAATTC TGTCTGTTTC	10800
TGTTTAACTT AATTAGTGAT GATAATATAA AGATGTATCA CTTGGTATAG AAGCTTTGGT	10860
ATTAAGTTTT TTATTAAGCC CATAAGGAAT ACCGATGGTT GGAGCAGCAG TTATAGCGTT	10920
CTTAGAAGGT ATAAATAGAA AAATAAGCTC ATTTTAAATC AAAGGATTGA TAAATCAGAA	10980

280

AGAAGGTGAT	TTTTTGCGAA	CATACGAAAA	TAAAGAAGAA	CTAAAGCTG	AGATAGAGAA	11040
AACATTGAG	AAATATATTT	TAGAATTTGA	TAATATTCCA	GAAAATTTAA	AAGATAAGAG	11100
AGCTGATGAA	GTTGACAGAA	CTCCAGCAGA	AAACCTTGCT	TATCAGGTTG	GTTGGACCAA	11160
CTTGCTTCTT	AAATGGGAAG	AAGATGAAAG	AAAGGGGCTT	CAAGTAAAAA	CACCATCGGA	11220
TAAATTTAAA	TGGAATCAAC	TTGGTGAATT	ATATCAGTGG	TTCACAGATA	CCTACGCTCA	11280
TTTATCTCTG	CAAGAGTTGA	AAGCAAAATT	AAATGAAAAT	ATTAATTCTA	TCTCTGCAAT	11340
GATTGATTCG	TTGAGTGAGG	AAGAATTATT	TGAACCGCAT	ATGAGAAAGT	GGGCTGATGA	11400
AGCGACTAAA	ACAGCGACTT	GGGAAGTGTA	TAAGTTTATT	CATGTAAATA	CGGTTGCACC	11460
TTTTGGAACT	TTCAGAACTA	AAATCAGAAA	ATGGAAGAAG	ATAGTATTAT	AAATTATATT	11520
TTTAACTTTA	AAAAATTTCA	TAAAAATGGT	TACCAAAGGC	GATAGAAGAA	AAACTATCGT	11580
CTTTTTCTTT	GCAAATTTTT	AAGAAGGGAG	GTGATCTTGC	ATGGACTTTG	AATATTTTTA	11640
TAACAGAGAA	GCGGAAAGAT	TTAACTTCTT	AAAAGTACCG	GAGATATTAG	TTGATAGAGA	11700
AGAATTTCCG	GGCTTATCAG	CAGAAGCAAT	TATCCTTTAT	TCCATACTTC	TTAAACAGAC	11760
AGGAATGTCA	TTTAAGAATA	ACTGGATAGA	CAAGGAAGGC	AGAGTATTTA	TCTATTTTAC	11820
TGTCGAAGAA	ATTATGAAAA	GAAGAAATAT	CTCAAAGCCA	ACTGCCATAA	AAACATTAGA	11880
TGAGCTTGAT	GTAAAAAAGG	AATAGGACTG	ATCGAAAGAG	TAAGGCTTGG	ACTTGGTAAG	11940
CCGAACATCA	TTTATGTTAA	AGACTTTATG	AGTATATTTT	AGGTAAAAGA	AAATGACTTA	12000
CAGAAGTCAA	AAAACCTAAC	TTCAGAAGTA	AAAGATTTTA	ACCTCAGAAG	TAAAGAAAAT	12060
GAACTTCAAG	AGGTTAAGAA	CCTTGACTCT	AACTATATAG	AGAATAATAA	GAGTAAGTAT	12120
AGTAAGAGAG	AATATAGTTT	TGGTGAAAAC	GGACTTGGAA	CATTTCAAAA	TGTGTTTTTA	12180
GCTGCTGAAG	ATATATCGGA	TTTACAAATC	ATAATGAACT	CACAGCTTGA	GAATTACATT	12240
AGACTTCCTG	CAAACTAGA	ATCCTAGTTC	ATGATTGATA	ATGCCAGCAA	TCAAATTCAT	12300
TCGTAATCCG	AAGCGTTTAC	GATGATTTTC	ATAGATTGTT	GAAAACATTT	TAAACGTTTT	12360
TACTTTGGCA	AAGATGTTCT	CAATCTTGCT	TCTCTCCTTG	GATAGCGCAT	GGTTACAGGC	12420
TTTATCTTCA	GCTGTTAGCG	GCTTGAGTTT	GCTGGATTTA	CGTGGAGTTT	GTAATTGAGG	12480
ATATATCTTC	ATGAGCCCTT	GATAACCACT	GTCAGACAAG	ATTTTACCAG	CTTGTCGGAT	12540
ATTTCTGCGA	CTCATTTTGA	ACAACTTCAT	ATCAGGACAA	TAGTTCACAG	CGATATCCAA	12600
AGAAACAATT	CTCCCTTGAC	TTGTGACAAT	CGCTTGAGCC	TTTATAGCGT	GAAATTTCTT	12660
TTTACCAGAA	TGATTGCTTA	ATTCTTTTTT	TAGGGCGATT	GATTTTTACT	TCCGTCGCAT	12720
CAATCATTAC	CGTGTCCTCA	GAACTGAGAG	GAGTTCTTGA	AATCGTAACA	CCACTTTGAA	12780

281

CAAGAGTTAC TTCAACCCAT TGGCTCCGAC GGATTAAGTT GCTTTCGTGA ATACCAAAAT	12840
CAGCCGCAAT TTGTCATAA GTTCGATATT CTCGCACATA TTGAAGAGTG GCCATAAGAA	12900
GGTCTTCTAG GCTTAATTTA GGTTCGTC CACCTTTTGC GTGTTTAAGT TGATAAGCTG	12960
TTTTTAATAC AGCTAATATC TCTTCAAAAG TCGTGCCTG AACACCAACA AGACGCTTAA	13020
ATCGTGCATC AGTTAGTTGT TTACTTGCTT CATCATTCAT AGAACTACTA TACCATATTT	13080
TGTTTCGCAG GAAGTCTATT GGAAAGTAAG AAATATTGAA GCTGAGGCTA TTAGAAGAAA	13140
TTGTGAGCGT GGTGCTATTT TTTGAGTAA AATAAAATAT CACGAAGATT CACAGTTTAA	13200
AGGAGATCAC TATGTTGAAT GTTATGCTGT TTTAGATAAT ACGGTTATAG CAAGAGATAG	13260
AATAACAGTC CCTATCGATC CGTTATGTGG AAAAGATTTT ATAGAGTAGC ATATAATTGA	13320
TTCTTAACTG GAATACTCAC TATCTCTTTA CATCAAGAAA ATGACTAAAC AGGGAAGTTT	13380
GCCTTCTTCC CTTTTTTTGT TATACTAGTA GAAGAAAAAA TTAGAAAGAT TTGTGGGTGT	13440
CAAACAGCCC AGTGGGGTGT TTTAATATGG ACTTAGGTCC CACCCAAAGA GGTATTAGTG	13500
TCGTGCTCA ATCTTATATC AATGTTATCG GTGCTGGTTT GGCAGGTTCT GAAGCAGCTT	13560
ACCAAATCGC AGAGCGTGGT ATTCCAGTTA AACTATATGA AATGCGTGGT GTCAAGTCTA	13620
CACCCAGCA TAAAACAGAC AATTTTGCTG AGTTGGTTTG TTCCAATTCT TTGCGTGGGG	13680
ATGCTTTGAC AAATGCAGTT GGTCTTCTCA AGGAAGAAAT GCGTCGCTTG GGTCTCTTGA	13740
TCTTGGAAATC TGCTGAGGCT ACACGTGTTT CTGCAGGTGG TGCCCTTGCA GTGGACCGTC	13800
ATGGTTTCTC TCAAATGGTG ACCGAAAAAG TTGCCAACCA CCCCTTGATT GAAGTGGTTC	13860
GTGATGAAAT TACAGAATTG CCGACAGATG TTATTACGGT TATCGCTACT GGTCTTTTGA	13920
GAAGTGAATG GTTGGGTGAA AAGATTGATG GTTCTTATGA TCGCTGCTGT TTTTATTTCT	13980
ACGATGCGGC AGCGCTATT ATCGATGTCA AACTATCGA TATGAGCAAG GTCTACCTCA	14040
AATCACGTTA TGATAAGGGA GAAGCGGCCT ACCTCAATGC CCTATGACC AAGCAAGAAT	14100
TTATGGATTT CCATGAAGCT TTGGTCAATG CAGAAGAAGC ACCGCTTAGT TCTTTTGAAA	14160
AAGAAAAGTA CTTTGAAGGA TGTATGCCTA TCGAAGTCAT GGCCAAACGT GGCATTAAAA	14220
CTATGCTTTA TGCCCTATG AAGCCAGTCG GTCTTGAGTA CCCAGACGAC TATACAGGAC	14280
CTCGTGATGG AGAATTTAAA ACACCTTATG CGGTTGTGCA ACTTCGTCAG GATAATGCAG	14340
CTGGTAGCCT CTACAATATT GTTGGTTTCC AGACCCACCT CAAATGGGGA GAACAAAAGC	14400
GTGTCTTCCA AATGATTCCG GGTCTTGAAA ATCGGGAGTT TGTCCGTTAT GGTGTGATGC	14460
ATCGCAATTC TTACATGGAT TCACCAAATC TTCTTGAGCA GACTTACCGT TCTAAGAAAC	14520



282

AACCAAATCT	CTTCTTTGCT	GGTCAAATGA	CGGGTGTGGA	AGGCTATGTT	GAGTCGGCGG	14580
CTTCAGGCTT	AGTTGCGGGA	ATTAACGCAG	CTCGTCTCTT	CAAGGAAGAA	AGCGAGGC7A	14640
TTTTCCCCGA	GACGACAGCG	ATTGGAAGCT	TAGCTCATT	CATTACCCAT	GCCGACAGCA	14700
AACATTTCCA	ACCAATGAAT	GTCAATTTTG	GGATCATCAA	GGAGTTGGAA	GGCGAGCGTA	14760
TCCGTGATAA	GAAGGCTCGT	TATGAAAAAA	TTGCAGAGCG	TGCCCTTGCC	GACTTAGAGG	14820
AATTTTGGAC	TGTCTAATTT	TTTGTAAAGA	ATTGCTCATG	ATACTATAAA	AATCTTAGAA	14880
ATTGTGATAA	AATAGGTAGG	ATGAAAGAAG	GAGACTGAAA	ATGGCGAATC	CCAAGTATAA	14940
ACGTATTTTA	ATCAAGTTAT	CAGGTGAAGC	CCTTGCCGGT	GAACGTGGCG	TAGGGATTGA	15000
TATCCAAACA	GTTCAAACAA	TCGAAAAGA	GATTCAAGAA	GTTCATAGCT	TAGGTATCGA	15060
AATTGCCCTT	GTTATCGGTG	GAGGAAATCT	CTGGCGTGGA	GAACCTGCAG	CAGAAGCAGG	15120
TATGGACCGT	GTTCAGGCAG	ATTACACAGG	AATGCTTGGG	ACTGTTATGA	ATGCTCTTGT	15180
GATGGCAGAT	TCATTGCAAC	AAGTTGGGGT	TGATACGCGT	GTACAAACAG	CTATTGCCAT	15240
GCAACAAGTG	GCAGAGCCTT	ATGTCCGTGG	ACGTGCCCTT	CGTCACCTTG	AAAAAGGCCG	15300
TATCGTTATC	TTTGGTGCTG	GAATTGGTTC	ACCTTACTTC	TCGACAGATA	CAACAGCGGC	15360
CCTTCGTGCA	GCTGAAATCG	AAGCAGATGC	CATCCTCATG	GCTAAAAATG	GTGTGATGG	15420
TGTTTACAAT	GCCGATCCTA	AGAAAGATAA	GACAGCTGTT	AAGTTTGAAG	AATTGACCCA	15480
CCGTGACGTT	ATCAATAAAG	GTCTTCGTAT	CATGGACTCA	ACAGCTTCAA	CCCTCTCAAT	15540
GGACAACGAC	ATTGACTTGG	TTGTATTCAA	CATGAACCAA	CCAGGCAACA	TCAAACGTGT	15600
CGTATTTGGT	GAAAATATCG	GAACAACAGT	TTCAAATAAT	ATCGAAGAAA	AGGAATAAGA	15660
AAGAATATGG	CTAACGCAAT	TATTGAAAAA	GCTAAAGAGA	GAATGACCCA	GTCTCACCAA	15720
TCACTTGCTC	GTGAATTTGG	TGGTATCCGT	GCTGGTGGTG	CCAATGCAAG	CTTGCTTGAC	15780
CGTGTACATG	TAGAATACTA	TGGAGTCGAA	ACTCCTCTTA	ACCAAATCGC	TTCAATTACG	15840
ATTCCAGAAG	CGCGTGTITT	GTTGGTAACA	CCATTTGACA	AGTCTTCATT	GAAAGACATC	15900
GAACGTGCCT	TGAACGCTTC	TGATATTGGT	ATCACACCGG	CTAATGACGG	TCTGTGATT	15960
CGCTTGGTTA	TCCCAGCTCT	TACAGAAGAA	ACTCGTCGTG	ACCTTGCTAA	AGAAGTGAAG	16020
AAGGTGCGCG	AAAATGCTAA	AGTGGCTGTC	CGCAATATCC	GTGCGGATGC	TATGGACGAA	16080
GCTAAGAAAC	GAGAAAAAGC	AAAAGAAATC	ACTGAAGACG	AATTGAAGAC	TCTTGAAAAA	16140
GACATTCAAA	AAGTAACAGA	CGATGCTGTT	AAACACATCG	ACGACATGAC	TGCTAACAAA	16200
GAGAAAGAAC	TTTTGGAAGT	CTAAAAATAA	ACAGAAAAAC	TCAGTTGGCA	TTGCTGGCTG	16260
AGTTTTATTTC	GAAAGAAGGA	AATATGAATA	CAAATCTTCG	AAGTTTTATC	GTTGGACTGA	16320

283

TCATCGATGA AAACGACCGT TTTTACTTTG TGCAAAAGGA TGGTCAAACC TATGCTCTTG	16380
CTAAGGAAGA AGGCCAACAT ACAGTAGGGG ATACGGTCAA AGGTTTTGCA TACACGGATA	15440
TGAAGCAAAA ACTCCGCTG ACAACCTTAG AAGTGACTGC CACTCAGGAC CAATTTGGTT	16500
GGGGACGTGT CACAGAGGTT CGTAAGGACT TGGGTGTCTT TGTGGATACA GGCCTTCCTG	16560
ACAAGGAAAT CGTTGTGTCA CTCGATATTC TCCCTGAGCT CAAGGAACTC TGGCCTAAGA	16620
AGGGCGACCA ACTCTACATC CGTCTTGAAG TGGATAAGAA AGACCGTATC TGGGGCCTCT	16680
TGGCTTATCA AGAAGACTTC CAACGTCTTG CTCGTCCTGC CTACAACAAC ATGCAGAACC	16740
AAACTGGCC AGCCATTGTT TACCGTCTCA AGCTGTCAGG AACTTTTGTT TACCTACCAG	16800
AAAATAATAT GCTTGGTTTT ATTCATCCTA GCGAGCGTTA CGCAGAGCCA CGTTTGGGGC	16860
AAGTATTAGA TGCGCGCGTT ATTGGTTTCC GTGAAGTGGA CCGCACTCTG AACCTCTCCC	16920
TCAAACCACG CTCCTTTGAA ATGTTGGAAG ACGATGCTCA GATGATTTTG ACTTATTTGG	16980
AAAGCAATGG CGGTTTCATG ACCTTAAATG ACAAGTCATC TCCAGACGAC ATCAAGGCAA	17040
CCTTTGGCAT TTCTAAAGGT CAGTTCAAGA AAGCTTTAGG TGGTCTTATG AAGGCTGGTA	17100
AAATCAAGCA GGACCAGTTT GGGACAGAGT TGATTTAGGG AGGCTTATGA GAAAATCATT	17160
TTACACTTGG CTCATGACCG AGCGCAATCC TAAAAGTAAC AGTCCCAAAG CAATTTTGGC	17220
AGACCTCGCT TTTGAAGAGT CAGCCTTTCC AAAACACACA GATGATTTTG ATGAGGTCAG	17280
TCGCTTTTGG GAGGAGCATG CCAGTTCTC TTTTAACCTA GGAGATTTTG ACAGCATTTG	17340
GCAGGAATAT CTAGAACACT AGCATTTATT CATTGGGTTT GGGCTAGTAA TTTCTCCATC	17400
CCTCTGCTAT AATAAAAAGA AATAAAAGGA TTAGAGAGGT TCTTTATTTG AAGGAACATT	17460
CAATAGACAT TCAACTGAGT CATCCAGATC ACCTGTTTCA TCTTTTGGT TCGAATGAAC	17520
GCCATCTTCG TTTGATGGAA GAAGAGCTTG ATGTTGTGAT TCATGCTCGT ACGGAGATTG	17580
TCCAGGTTTT GGGAGAAGAG TCTGCCTGTG AGGAAGCCCC TCAAGTTATT CAGGCTTTGA	17640
TGGTCTTGGT AAATCGTGGG ATGACCGTTG GTACGCCAGA TGTAGTCACT GCGATTAGCA	17700
TGGTCAAAAA TGATGAAATT GACAAGTTTG TCGCCCTTTA CGAAGAAGAA ATTATCAAGG	17760
ATAATACTGG GAAACCTATC CGTGTCAAAA CCCTAGGGCA AAAGCTTTAT GTGGACAGTC	17820
TCAAACAGCA TGATGTGACC TTTGGAATTG GGCCAGCAGG TACAGGGAAG ACCTTCCTTG	17880
CAGTGACCTT GGCAGTGAAT GCCCTTAAAC GTGGGCAAGT CAAGCGAATT ATCCTAACTC	17940
GTCCAGCGGT GGAAGCGGGA GAGAGTCTTG GATTTCTTCC GGGTGATCTT AAGGAGAAGG	18000
TGGATCCTTA CCTTCGTCCT GTTTACGATG CCTTGATCA AATTCTTGGG AAAGACCAAA	18060

284

CGACTCGTCT CATGGAGCGT GAAATTATCG AAATTGCGCC CCTTGCCTAT ATGCGTGGCC	18120
GGACCTTGGA TGATGCCCTT GTCATTCTCG ATGAGGCGCA AAACACGACC ATCATGCAGA	18180
TGAAGATGTT CTTGACGCGT TTAGGTTTTT ATTCTAAGAT GATTGTCAAT GGAGATATTA	18240
GTCAGATTGA CCTGCCACGT AATGTCAAGT CCGGTTTGAT TGATGCTCAA GAGAAACTCA	18300
AGAACATCCA TCAGATTGAC TTTGTTCAAT TTTTCAGCCAA GGATGTGGTT CGCCATCCTG	18360
TTGTCGCTCA GATTATCCGA GCCTATGAAT ATTCTACTGA AGTTGCACAC GACTGATTTT	18420
GAGGAAGTTC GCCTGCAAAA GAATAGACTT GTTCGGTAAC TGTA AAAAGT GTTATACTAT	18480
TTTTATGGAA ACAGTATACG ACAAAGCACA AAAACTTAAC TCAAAAACT TCAAACTATT	18540
GATTGGTGTC AAAAAGGAAA CCTTTCAACT CATGCTAGAA CACCTGAATT CAGCCTATCA	18600
GATTCAGCAC CGAAAAGGTG GACGTCCACG TAGTCTGCCC ATGGAAGACC AGCTCATTAT	18660
GACCCCTCCGT TACTTGCGAT ATTATCCAC TCAGCGTCTG CTGGCCTTTG ATTTTGCGT	18720
CGGTGTAGCT ACGGTAAATG CCATCATCAC TTGGGTGGAG GATACACTTC GTGCGTCAGG	18780
TAGCTTTGAT TTGGACCATT TAGAAGCCCC GAGTGCTGCT GTGGCTATTG ACGTGACCGA	18840
AAGTCCGATT CAGCGTCCAA ACAAACCAA AGCAAAAATT ATTCTGGTAA AAAGAAACGA	18900
CACACCTTAA AACTCAAAT TATGCTGGAT TTGACGACAC ATAAAGTCTG TCAAAATGGCC	18960
TTTTCTGACG GACATACGCA TGATTTTACT CTCTTCAAAG AAAGTATTGG ACAAAGTTTG	19020
CCTGAAACGA CGCTTGCCTT TGTGACCTA GGTATTATTAG GCATCTTGAA ATTTTCATGAG	19080
AATACTTTCA TTCCTGCTAA AAATTCCAAA AATCGCCGCC TGAGTGAGGA TGATAAGCAG	19140
TTAAATAAAG AGATGTCAGC GATACGAATT GAAATTGAAC ATTTTAACGC TAAATTCAAG	19200
ACCTTCCAAA TCATGTCAGT CCCTTATCGT AACCGCAGAA AACGTTTCGA GTTACGGGCG	19260
GAATTAATTT GTGCCATCAT CAATTATGAA GTGAAGTAGA TTCCGAACAA GTCTAATATA	19320
CTTTTGAGAG AGGAAAATCC AGTTGTATAG GCTAAAGGTT TTATCCAAAG GTCTGAGACA	19380
ACGATTAGGC ACGATGGAAA GAACTTTTAT GTGGCTGATG ACGATCAGTG CATCTTCCTG	19440
TGTCATAATC ACAGGGCACA AGAAAGTAGG AATTTGAAAA GATGATTGAC CAACTATCTA	19500
AGTATTACAG TTGTAGGATA CTAAGTAAA AGGATATTCC AAGTATTTTA TCTTTATATG	19560
AAAGTAATCC TCTGTATTTT CAGCATTGTC CACCAGAGCC AAATTTTGCA ACTGTAAAAG	19620
AGGACATGCT TTGTCTACCT GAAGGTAAA GCTAAGGCTGA TAAGTTTTTT GTTGGATTTT	19680
GGAATGGATC TGACCTTG TGCTGTTATGG ATTTTCTCTA TGCATATCCT GATGAGGAGA	19740
CTGTTTTTAT TGGTTTGTAT ATGGTTGATC AAGCCTATCA GAGAAAAGGG ATTGGTAGTC	19800
ATATTGTGAC AGAAGCACTA GCTTATTTTG CTAAGAACTT TCGAAAGGCA CGTTTGGCTT	19860

285

ATGTTAAGGG AAATCCGCAA TCTCAGCATT TTTGGGAAAA GCAGGGCTTT AAATCAATTG	19920
GATGCGAGGT TAAGCAAGAA CTTTATACGG TTGTTATCGC TGAACAGAGC CTAGAAGATT	19980
AGAAATGGCA TCAAGTAAGA ACTATTTGGA ATTTGTTTTG GAACAATTAT CAGGATTAGA	20040
TGATGTGACT TACCGTTCCA TGATGGGGGA GTATATTCTT TACTTCCGCG GCAAGATTAT	20100
TGGCGGCATT TATGACGATC GCTTTT TAGT TAAACCCGTC CAAGCAGTCT TAGATAAGAT	20160
TGACCAATCT TCTTTTGAGT TTCCATACAA AGGTGCCAAA GAAATGATTT GAGTGGAAGA	20220
ACTTGATAAT AAGATGTTTC TATAAGACCT AATTTTAGCT ATGTATAACC AACTGCCAAC	20280
GCCCAAACCT AAAAAGAAAA AGCAAGGGTG AACGAAGTAA AAAAGAAGTC TGCTAAGGCC	20340
CTGTCTTTGC ACGGGTAAAA TTTTATATAT AAAAAGAAGC TGGGACTAAA GAGCTCAGCT	20400
TCCTTTGGTT TATATAATTG TCATTACAAG ACGAAGTGGT TGGGCGAAAC TCTGTTGACT	20460
TTATTCAATT TAGAGTTTCT TATGCACAAT TGACTCTGGA ACGAAAGTCT CCAGTTGCAA	20520
AGTATACAGT ACAATAAACC AACGATGTAA TAGCTGATGA CACAAAGCAC AGTGGGTAGG	20580
ACTTCCGAAG TCACCCTTTT CTTTTCAAAA TTTATACTAA ATCATTGATA TCAGTGACT	20640
CACGATTAAG TCCTTGAGCA ACTGGTAGGT TAGTCAAGTA ACCTTGATAA GTAGTCACAC	20700
CTTGACGCAA GCCTTCATCT TCAGAGATTG CTGTGCGAA TCCTTTGCCA GCCAAAGCTT	20760
CGATATAAGG AAGAGTGACA TTGGTTAGGG CGATGGTTGA AGTGCGAGCA ACCGCACCAG	20820
GGATATTGGC AACGGCATAG TGGAGAACAC CGTGTTTTTC ATAGACGGCT TCATCGTGCG	20880
TTGTCACACG GTCAGCTGTT TCGATAACGC CACCTTGGTC AACAGCAACG TCAACGATAC	20940
AGAGCCTGGA CGCATTTGTT TGACCATCTC ATCTGTCACC AATTCCGGTG CTTTTCACC	21000
AGGGATGAGA ATGGCTCCAA TCACCACATC AGCATCTCTC ACACTTGCTT CAATGTTGAA	21060
TGAATTAGAC ATAAGAGTTT GAATTTGACT TCCAAAGACT TCTTCTAGAA CTGAGAGACG	21120
CTTGGAAC TAATCTAAAA TAGTCACTTG AGCACCAAGA CCAAGGGCGA TGCGGGCAGC	21180
ATGTGTACCG ACGACACCAC CACCGATGAT AGTTACTTTT CTTTTTGAA CACCTGGTAC	21240
ACCACCAAGT AGAACACCAG AGCCACCAGC TTGCTTAGTA AGGAAGTGAG CTCCGATTTG	21300
AACAGCCATA CGACCTGCAA CCTCACTCAT AGGAACGAGG AGCGGTAGTT GTCCTTGATT	21360
GTCACGAACA GTTTCAGTTG TTTTGTCTGT TAACATAGCA TCTGCTAATT CTGGAGCAGC	21420
GGCCATGTGC AAGTAGGTGA AGAGAAGAAG ATCGTCGCGC AAGTAACCGT ATTCAGAACT	21480
TAAAGATTCT TTTACTTTCA CAACCAACTC TGCTGCCCAA GCTTCACCAG CAGTAGCGAC	21540
AATCTCAGCT CCTTGCTTTT GATAGTCAGC ATCAGTAAAG CCAGAACCGA GACCAGCATT	21600

286

TGTTTCGATA	AGGACACGAT	GACCACGACT	AACTAAGCTA	TGAACACCTG	CAGGTGTGAG	21560
GGCGACACGG	TTTTCGTTAT	TTTTAATTTT	TTTTGGGATT	CCGATTAACA	TTGAGATAAC	21720
CTACCTTTCA	ATTGACGGTC	TTGTTTTGGT	TGTCACATTC	CAGTTCATAA	ATCAAAAATG	21780
TGACGGTTTC	ATTGTATATG	AAACCGCTTC	AAAAATCAAG	AAAAACTTGT	CATCCAAATT	21840
TTTTTATGCT	AGACTAGTGA	AAATCAAGCT	CTAATGGAGG	GAAAAGTATG	GAATCAATAT	21900
TTGTGAAATT	TGCCCAGTAT	CCGTCTATAG	AAACGGAGCG	TTTATTGCTC	AGACCTGTAA	21960
CTTTGGATGA	TGCGGAACAA	TGTTTGACTA	TGCCTCGGAC	AAGGGTAATA	CACGTTACAC	22020
TTTTCCAAACC	AATCAAAGCT	TGGAAGAAAC	CAAGAATAAC	ATTGCTCAGT	TCTACTTGGC	22080
TAATCCCTTG	GGACGTTGGG	GAATAGAACT	AAAAAGCAAT	GGTCAGTTTA	TTGGAACCAT	22140
TGACTTGCAC	AAGATTGATT	CTGTTCTTAA	GAAGGCAGCT	ATTGGCTACA	TTATCAATAA	22200
AAAGTATTGG	AATCAAGGAT	TAACGACAGA	AGCCAATCGT	GCTGTGATTG	AGCTAGCTTT	22260
TGAGAAGATA	GGGATGAATA	AGTTGACTGC	CCTTCACGAT	AAGGCTAATC	CCGCGTCAGG	22320
AAAGGTCATG	GAGAAATCAG	GCATGCGTTT	TTCCCATGCA	GAACCATATG	CTTGTATGGA	22380
CCAGCATGAA	AAAGGCCGAA	TCGTGACAAG	AGTTCATTAT	GTCTTGACCA	AGGAAGACTA	22440
TTTTGCAAAT	AAATAAGCAG	TTGAAAAGAA	ATTTTTTCGAC	TGTTTTTTCT	TCCTCTTACC	22500
AATAATCTAA	GAGAGGAGAA	AATATGGAAG	CAATTATCGA	GAAAATCAAA	GAGTATAAAA	22560
TCATCGTCAT	CTGTACTGGT	CTGGGCTTGC	TTGTAGGAGG	ATTTTTCTTG	CTAAAACCAG	22620
CTCCACAAAC	ACCTGTCAAA	GAGACGAATT	TGCAGGCTGA	AGTTGCAGCT	GTTCCTCAAGG	22680
ACTCATCGAC	CGAAAAGGAA	GTGAAGAAGG	AAGAAAAGGA	AGAACCCTTT	GAACAAGATC	22740
TAATCACAGT	AGATGTCAAA	GGTGTCTGTA	AATCGCCAGG	GATTTATGAC	TTGCCTGTAG	22800
GTAGTCGAGT	CAATGATGCT	GTTCAGAAGG	CTGGTGGCTT	GACAGAGCAA	GCAGACAGCA	22860
AGTCGCTCAA	TCTAGCTCAG	AAAGTTAGTG	ATGAGGCTCT	GGTTTACGTT	CCTACTAAGG	22920
GAGAAGAAGC	AGTTAGTCAA	CAGACTGGTT	CGGGGACAGC	TTCTTCAACA	AGCAAGGAAA	22980
AGAAGGTCAA	TCTCAACAAG	GCCAGTCTGG	AAGAACTCAA	GCAGGTCAAG	GGACTGGGAG	23040
GAAAACGAGC	TCAGGACATT	ATTGACCATC	GTGAGGCAAA	TGGCAAGTTC	AAGTCAGTAG	23100
ACGAGCTCAA	GAAGGTCTCT	GGCATTGGTG	GCAAAACAAT	AGAAAAGCTT	AAAGACTATG	23160
TTACAGTGGA	TTAAGAATTT	CTCTATTCCC	CTAATTTACC	TGAGTTTTCT	ATTACTTTGG	23220
CTTTATTACG	CTATTTTCTC	AGCATCTTAT	CTTGCTTTGT	TGGGCTTTGT	TTTTCTGCTA	23280
GTCTGTCTCT	TTATCCAATT	TCCGTGGAAA	TCTGCTGGTA	AAGTTCTAAT	AATTTGCGGA	23340
ATCTTTGGAT	TTTGTTTGT	TTTTCAAAAT	TGGCAACAGA	GTCAAGCGAG	TCAAAATCTG	23400

287

GCGGATTCTG TTGAAAGGGT ACGGATTTTG CCTGATACTA TTAAGGTTAA TGGTGATAGT	23460
CTATCCTTTC GTGGCAAGTC TAACGGTCGT GCTTTCCAAG TCTATTATAA ACTCCAGTCC	23520
GAGGAGGAGA AAGAAGCCTT TCAAGCTTTA ACTGACCTGC ATGAGATAGG ACTAGAAGGG	23580
AAGCTTTCGG AGCCAGAAGG GCAGAGAAAT TTTGGTGGCT TTAATTACCA AGCCTATCTG	23640
AAGACTCAGG GAATTTACCA GACTCTCAAT ATCAAAACAA TCCAGTCACT TCAAAAGATT	23700
GGCAGTTGGG ATATAGGAGA AAACCTGTCC AGTTTACGTC GAAAGGCTGT GGTTTGGATT	23760
AAGACGCACT TTCCAGACCC TATGGGCAAT TACATGACAG GACTCTTGCT GGGACATCTG	23820
GACACCGACT TTGAGGAGAT GAATGAGCTT TATTCCAGTC TAGGAATTAT CCACCTCTTT	23880
GCCCTATCTG GCATGCAGGT AGGTTTTTTC ATGAATGGAT TTAAGAACT TCTCTTGCGA	23940
TTGGGCTTGA CCCAAGAAAA GTTGAAATGG CTGACTTATC CTTTTCCCT TATCTATGCG	24000
GGACTAACTG GATTTTCAGC ATCGGTTATT CGCAGTCTCT TGCAAAAGCT ACTGGCTCAA	24060
CATGGGGTTA AGGGCTTGA TAATTTTGCC TTGACGGTGC TTGTCCTCTT TATTGTCATG	24120
CCAAACTTTT TCTTGACAGC AGGAGGAGTC TTGTCCTGCG CTTATGCTTT TATCCTGACC	24180
ATGACCAGCA AAGAAGGGGA GGGGCTCAAG GCTGTTACTA GTGAAAGTCT AGTCATCTCC	24240
TTGGGCATAT TGCCCATCTT ATCCTTCTAT TTGCGGAAT TTCAACCTTG GTCTATCCTT	24300
TTGACCTTTG TCTTTTCTTT TCTTTTGGAC TTGGTCTTCT TACCGCTCTT GTCTATCTTA	24360
TTTGTCTTTT CTTTCTCTA TCCAGTCATT CAGCTGAACT TTATCTTTGA ATGGTTAGAG	24420
GGCATTATTC GCTTGGTCTC GCAGGTGGCA AGGAGACCAC TTGTCTTTGG TCAACCCAAC	24480
GCATGGCTTT TAATCTTATT GTTAATTTCC TTGGCTTTGG TCTATGATTT GAGGAAAAAC	24540
ATTAAAGGAT TAACAGTATT GAGTTTATTG ATTACAGGTC TCTTTTTCCT TACCAAGTAT	24600
CCACTGGAAA ATGAAATCAC CATGCTGGAT GTGGGGCAAG GAGAAAGTAT TTTCTACGGG	24660
ATGTAACCTG GAAAACCATT CTCATAGATG TAGGTGGTAA GGCAGAATCT TATAAGAAAA	24720
TCAAAAAATG GCAAGAAAAG ATGACGACCA GCAATGCCCA GCGAACCTTG ATTCCCTATC	24780
TCAAAAGTCS AGGAGTAGCT AAGATTGACC AGCTAATTTT GACTAACACG GACAAGGAGC	24840
ATGTTGGAGA TTGTTCAGAG ATGACCAAGG CTTTCCATGT AGGGGAGATT CTAGTATCAA	24900
AAGACAGTCT GAAACAGAAG GAATTTGTGG CAGAACTACA GGCGACTCAA ACAAAGGTGC	24960
GTAGTATGAT AGTAGGGGAG AACTTGCCCA TTTTGGAAAG TCAGTTAGAA GTTCTATCTC	25020
CAAGGAAAAAT GGGAGATGGA GGACACGATG ATACCCTAGT TCTGTATGGG AAATTCTTGG	25080
ATAAGCAATT TCTTTCACG GGAAATTTGG AGGAGAAAGG AGAGAAGGAC TTGCTGAAGC	25140

288

ACTATCCAGA CTTGAAAGTA AATGTTTTGA AAGCTAGCCA ACATGGCAAT AAAAAATCAT	25200
CAAGTCCAGC CTTTCTAGAA AAACCTCAAAC CAGAGCTTAC TCTTATCTCA GTTGGAAAGA	25260
GCAATCGAAT GAAACTCCCC CATCAGGAAA CATTGACACG ACTGGAAGGT ATCAATAGCA	25320
AAGTTTATCG AACTGACCAG CAAGGAGCTA TACGTTTTAA GGGGTTGGAT AGTTGGAAAA	25380
TCGAAAGTGT TCGATAGGAA GGATAAATGT TGTAGATTAG TGAAATAAAC TAAAAATTTG	25440
TTGCATAATA ATGATAAAAA TGGTATAATG AAAACGTATT CAATATTGAG GATATAAAAT	25500
CATTAAAAAT CAGCAAAAGT TGTTTTATTA GTTAGTTTAT AATCTATTGG TCTTCTTCAG	25560
TCCAGTGTAT CTGCTGTGAC AGTCACTAAA AGTTACAAGT ATGATTGGAA TACGGTTTGG	25620
GAATATAGTA CCAACTATCA CGACCATCAG TATGCTTGGA TTCCGTCATC GTCTCGTTAT	25680
GACAGCTATT CTGAGTATAA AGTTGGCGGA GGCTGGAACT ACGCTCGTTA TGAGGTCATA	25740
AACTATTACA GCGGAGGCTA TTAATTCTTA AAGAGTGAGA AAAAGGAGGG CTAGATATGT	25800
TGCAGCTTAC TCATGTGACC TAAAAACGC GACAAGTCAT CTTGCAAGAT GTGGATTTCA	25860
CCTTTAAAAA GGGTAGGGTT TATGGTCTTC TTGCTATCAA TGGCTCTGGA AAGACGACCC	25920
TGTTCCGTGC CATTAGCAAT TTAATTCCCA TAAGTAGTGG AAATATCGCA GCCCCTCCTT	25980
CTTTATTTTA TTATGAGAGT ATTGAATGGC TGGATGGAAA CTTAAGTGGG ATGGACTACC	26040
TTCGTCTTAT CAAAAACATC TCGAAGTCAG GTCTGAACTT GAGGGATGAA ATCGCCTATT	26100
GGGAAATGTC TGAATATATC AGTCTTCCCA TTCGCAAGTA TTCCTTAGGC ATGAAGCAAC	26160
GCTTGGTGAT TGCCATGTAT TTCCTCACTC AGGCCAAATG CTGGCTCATG GATGAGATTA	26220
CAAATGGCTT AGATGAGTAT TATCGACAGA AGTTTTTTGA TAGGCTAGCA CAAATCGATA	26280
GACAAGAACA GCTGGTCTCT TTAAGTTCCC ACTATAAGGA AGAGTTGGTT GATGTCTGCC	26340
ATAGAGTAGT AACCATTCAT CAGGGGCAGA TAGAAGAGGT TTAGTTTATG AAAGATGTTA	26400
GTCTATTTTT ATTGAAAAAA GTTTTCAAAA GCCGCTTAAA CTGGATTGTC TTAGCTTTAT	26460
TTGTATCTGT ACTCGGTGTT ACCTTTTATT TAAATAGTCA GACTGCAAAC TCACACAGCT	26520
TGGAGAGCAG GTTGGAAAGT CGCATTGCAG CCAACGAGAG GGCTATCAAT GAAAATGAAG	26580
AGAAACTCTC CCAAATGTCT GATACCAGCT CGGAGGAATA CCAGTTTGCT AAAAAATAAT	26640
TAGACGTGCA AAAAAATCTT TTGACGCGAA AGACAGAAAT TCTGACTTTA TTAAAGAAG	26700
GGCGCTGGAA AGAAGCCTAC TATTTGCACT GGCAAGATGA AGAGAAGAAT TATGAATTTG	26760
TATCAAATGA CCCGACTGCT AGCCCTGGCT TAAAAATGGG GGTGACCGC GAACGGAAGA	26820
TTTACCAAGC CCTGTATCCC TTGAACATAA AAGCACATAC TTTGGAGTTT CCGACCCACG	26880
GGATTGATCA GATTGTCTGG ATTTTAGAGG TTATCATCCC AAGTTTGTTC GTGGTTGCTA	26940

289

TTATTTTAT GCTAACACAA CTATTTGCAG AAAGATATCA AAATCATCTG GACACAGCTC	27000
ACTTATATCC TGTTCACAAA GTGACATTTG CAATATCCTC TCTTGGAGTT GGAGTGGGAT	27060
ATGTAACGTG GCTGTTTATC GGAATCTGTG GCTTTTCTTT TCTAGTGGGA AGTCTGATAA	27120
GTGGTTTTGG ACAGTTAGAT TATCCCTACC CAATTTATAG CTTAGTGAAT CAAGAAGTAA	27180
CTATTGGGAA AATACAAGAT GTATTATTTT CTGGCTTGCT CTTAGCTTTC TTAGCCTTTA	27240
TCGTCAATTCT GGAAGTTGTG TACTTGATTG CTTACTTTTT CAAGCAAAAA ATGCCTGTCC	27300
TCTTTCTTTC ACTCATTGGG ATTGTTGGCT TATTGTTTGG TATCCAAACC ATTCAGCCTC	27360
TTCAAAGGAT TGCACATCTG ATCCCTTTA CTTACTTGCG TTCAGTGGAG ATTTATCTG	27420
GAAGATTACC TAAGCAGATT GATAATGTCG ATCTAAATG GAGCATGGGA ATGGTCTTAC	27480
TTCCTTGCCT GATTATCTTT TTGCTATTGG GAATTCTATT TATTGAAAGA TGGGGAAGTT	27540
CACAGAAAAA AGAATTTTTT AATAGATTCT AGCTTTCCTA TAGGTAGGGA AAATAAGTAA	27600
AACTAACAT AGAGAGGGAA TCAACTTGAT TCTCTCTTTT TGATTGAAA ACCAAACCAA	27660
AATACAAACA CAACTTTTC AAAAAATAAC TTTTATCTT GACAAGAGCT AGAAAACTTG	27720
GTATCATATA AAAGTTGAGA AAAGCAGAAG TGAGAGCTTC TCGCCTTGTG ACATTAAGTT	27780
GCCTGGCCCT ACGGATGAAA AGTTTCGAAG AAACGCTATC ATAACGTGGG GGCTTGATA	27840
TTTACAAGTC CGCTATTGTT TTTCTCTAAT AAAACAAAAG AGGTGAAAAC CATAGCAAAG	27900
CAAGACTTAT TCATCAATGA TGAGATTCGT GTACGTGAAG TTCGCTTGAT TGGTCTTGAA	27960
GGAGAACAGC TAGGTATCAA GCCACTCAGT GAAGCGCAAG CTTTGGCTGA TAACGCTAAT	28020
GTGACCTAG TATTGATTCA ACCCCAAGCC AAACCGCCTG TTGCAAAAAT TATGGACTAC	28080
GGAAGTTCA AATTGAGTA GGAAGAGAG GAAAGAGAG AAGCTAAAA AAAAACTT	28140
GTTACTGTGA AAGAAGTTGG TCTAAGTCCG G	28171

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7147 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

CCGCTCAACT TTTGCAATCA AGGCTAAGTA GACAGCAGCA AATTTCATAT TGTATAATTT	60
CTGACTCATA CTTCTCTCTT TCTATGTGTA CTAGTATAAA TAAGAAAAAG AAGGCCGTCA	120



290

AGCCTTCTTT TGATTATTC TTCTGCTTCA TCTTCTGTAA ATTGACTATT GTACAAGTCA	180
GCGTAGAAGC CACCTTGCGC CATCAGTTCC TCATAGTTGC CTTGCTCGAT GATATTTCCA	240
TCTTTCATGA CCAAGATCAA GTCTGCATTT CGGATGGTTG ACAAGCGGTG GGCAATGACA	300
AAGGATGTGC GTCCTTCCAT CAAACGGTCC ATGGCTTTTT GGATCAATTC CTCTGTCCGT	360
GTGTCAACAG AAGAAGTCGC CTCATCCAAA ATCAAAAGCG GTGCATCCTT AAGAAGGGCA	420
CGAGCAATAG TCAATAGTTG TTTTGTCTT ACAGACAAGG TCACGGTGTC ATCCAAGATG	480
GTATCATAGC CATCTGGCAA GGTGATAATA AAGTGGTGAA TTCCACAGC CTTACTAGCT	540
TCCATCATTC GTTCATCACT AATCCCTATT TGATTATAGA TGAGATTGTC TCGAATAGTT	600
CCTTCAAAGA GCCAGGTATC CTGCAAGACC ATTGAAAAGG CATCATGCAC TTCTGAACGC	660
GTCATAGCCT TGGTATCCAC ACCATCAATG CGAATACTTC CCTTATCAAT CTCATAGAAT	720
TTTATCAAAA GATTGACAAT GGTGTCTTA CCAGCCCCAG TCGGCCAAC AATGGCAACC	780
TTTTGACCAG CATGAGCTGT CGCAGAGAAG TCATAGTCTT GAACATGAC ACCGTCCACC	840
AGAATTTCTC CTGCTGACAC GTCGTAGAAA CGTGGAAATCA GATTGACCAG AGTTGATTTA	900
CCAGAACCTG TTGACCCAAT AAAGGCCACT GTTTGACCAG TTTCTGCTTT AAAGCTAACA	960
TGTTCAATAA CTGCCTCCGA ATTTGCCGCA TAGCGGAAGG TCACATCCTT AAATCSACC	1020
TGACCTTTGA AGTTTTCATC AGTCAGCTGC ACTTGAACAG CGTTTTGGAT AGAAGAATGC	1080
AAATCTAAAA CTTGATTAAT CCGCTTAGCA GAGACCATAG TTCGGGGAAG AACGATGAAG	1140
AGTGCTCCCA TGAGAAGGAA GCCCATGACA ACCTACATGG CATAAGACAT GAAAACAATC	1200
ATGTCACTAA AGAGAGGCAG ACGCGCTATC GGAGCAGCGT CGTTAATCAC ATAGGCCCCA	1260
ATCCAGTAAA TCGCCACACT CAAACCACTT GAAATCCCCA TCATCATAGG ATTCAAAATA	1320
GCCATAAGAC GGTTGACAAA CAAATTCAAA CGGGTCAATT CATCATTTAC TGCTGCAAAT	1380
TTTTCATTTT GATAATCCTC TGCATTGTAG GCACGAACGA CACGAATACC TGTTAACTC	1440
TCACGAGTGA TACTGTTTCTG TTTATCTGTC AGCCCCGAA TCAAGGACTG TTTTGAAAG	1500
GCTAGCGTCA TCAAAACGGT CGTCATCAGG ACGTTGATAA TCACTGCCAC AAGTACGGCC	1560
CAGAGCCAGT ATTCCTGAATG ACCTAAAATC TTCCCAATAG CCCAGATAGC CATAATTGAA	1620
CCACGCGTTA CCACCTGCAA GCCCATAGTA ATCAACATTT GAACTTGAGT AATGTCATTG	1680
GTAGTACGCG TCAAGAGGCT AGGAATTGAA AATTTCTTAA TCTCTGCTG CGAGTAATCC	1740
AAAACCGGT TAAAAATATC ACTTCTCAGC CTACTAGTAT AAGAAGCCGC CACTCGGGAT	1800
GCAAAAAATC CAACTGCAAC TACGGACAAG AAGGCAAGAA AGGACATTCC CATCATCATG	1860
CTTGCCGACT GCCACAATC ATCTAAATTA GTTCTTGAC TACCTAGCAA ATCCGTAATT	1920

291

TTCGAGATAT AGGTCGGCAC TTCCAACCTCT AGATAGACCG AAAAGCAAGT AAAGAGAATG	1980
GCTAGTAAAA TCATCCCCCA TTCTTTTCTA CTAATTCTTT TGGCTAATTT CTTTATTCTC	2040
TCCTCCTATT CCCTTGATAT TTTGCCTGTA GTTGACCGAG AACCTTCTCA AAAATCAGTA	2100
ATTCATCTTC ATCAATGTCT TCCATCAACT GCTTGTCTAT GCGTTCAAAA AAAGCCTTAA	2160
CCTGTTGCAT CTGAGAACGT GCTTGTCCG TCAGACGAAC AAAGTTAGCC CGCTTATCAA	2220
CAGGACTCGC CTCCAATTCC ACCAAACCAT TTTGCACTAT ACGCTTAACC AGATTACTAG	2280
CAACAGGCTT GGTAAATATG AGTTCCTGCT CGATATCTTT AATCAAGACC AAGTCTTGGT	2340
TTTTCTCGCG ATTATCCAAA AAACGCACAA CCTGACCTTG CGGCCACCC ATAAATTCAA	2400
TGCCGCAACG TTTGGCTTCC TTTTGCACCA TCAGGTGAAT TTGATGACCA AAACGCTTAA	2460
AGACTAACAT CGGTTTATCC ATAATCTCCC CTTTCTAAAT AAAAATAGTT CTCTGGAGAA	2520
TAATTAAAT TCTATGAGAA CTATTTTCTT GATTAAAAAA ATCCCAAGTG ATTTTCTCAC	2580
TTAGGATCAT GTTCTATAGG TTAAATTAAA ACCCATCTAC GTTCGTATAA ATCTTTTGGG	2640
CGCTTTCGTC GTCTTCAAGA ACGCTGTAAA GTTTTCAAAA GGTTCGAAGG TCTTCGCTG	2700
ACAATTCCAC TTCTGACTGA GGAATCATTT CCAATTCAGT CACTTGAAT TCTTCAATAC	2760
CAGACTCAGC GAGGGCAACG ATAGCCTTGT GAAGGTCAGT TGGCGCTGTG TAAACTGTGA	2820
TTGTACCTTC TTGTGCTTCT ACGTCATCCA CATCCACATC CGCTTCGAGC AATTGCTCAA	2880
AGACTGCGTC CGCATCTTCA CCTCCAAATA CAATAACACC TTTGTTGTCA AAGAGGTAAG	2940
AAACAGAACC TGAAGCGCCC ATGTTTCCGC CGTTTTTACC AAAGGCTGCA CGGACATTGG	3000
CTGCTGTACG GTTGACGTTA GAAGTCAAAG TATCCACAAT TAGCATAGAG CCATTTGGCC	3060
CAAAAGCTTC GTAACGTCTT TCTGTAAAGG TTTCTCTCTG GTTCTCTTTC GCTTTATCAA	3120
TCGCTTTATC GATAATGTGT TTTGGCACTT GGGCTTGTTC AGCACGGTCG ATAACGAATT	3180
TCAAAGCTGA GTTTGATTCT GGATCTGGAT CACCTTTTTT AGCTGCTACA TAGATTTCTA	3240
CACCAAATTT TGCATATACT TTAGAGTTAG CTCCATCTTT AGCCGTTTTT TGGCTACGA	3300
TATTGGCCCA TTTACGTCCC ATTAGGAATC TCCTTTTTTC ACATTTTAAT CTTTCTTATT	3360
ATAACACAAG TTTTTTTGAT TTCACTAGA GGAAATGGAT TTTATTAGCA AATCAAGCTA	3420
GGATAGCACT TTACCTGCTA AGATGGTCTT GCCTTTCTAT CTTTATCAAC AGGCACTCAT	3480
CCACATTCAA AAAACAAACT AGACCATTAT CTGCAAATAG AAAGTTTCAG CCAAGTTTGA	3540
CAAAGTCAGC TCAAATTACT GTTTGAAGTT TGATAGATATA AGCGACAAAA ACAATCATAC	3600
TGCACCTTTT GTTGACAGTC TACTCCAGAC ATATCATAGT TCAAGTAAAT ACTTTGAAAT	3660

292

TCAACAGTTC TTATAGGCGC TATTGTATTC TAAGAAATCA ATAGAAAGT TTCTAAGCAA	3720
ACCTCTAATA CTCAATAAAA ATCAAAGAGC AAAC TAGAAA GCTAGCCTCA GGTGCTCAA	3780
AACACTGTTT TGAGGTGCG GATGGGGCTG ACATGGTTTG AAGAGATTTT CGAAGAGTAT	3840
AATTTACGTG TTCCCAAGAT GGAGAAGTTA GACTAGTACA CTGGCACTTC TAAAACATTG	3900
CTAGCAATTG ATTTGTTCAT ATTTAATTTT ATTTTTTCCA TAAATGGGTA TTAGATATAA	3960
ACAGCAAAAT ATTTCCGATA CGTGTCTGTC TTGAATTTCC AATCATCTAA AACAAGTAAA	4020
GGATAATCAA TCCCCTGTAT ATCAAGGAAT TGGCTACCCT TTTTACTTTT TTACACATTC	4080
TGTTTGATAG ATTCATTTTA ACATCACGAG CATACTCCAA TGGAAATCGC TAGGCAAGAG	4140
ATAAACTTTC AGATATCCGC AGAGAGATCA TCGCCTCTTT TTGTCCGAAG CATTCTCCTC	4200
TCCTAGTCAT TTTCTACCTT ATCTTCTACC TGAGGATAGA GAGTTGTTC CCAAATAGAA	4260
ATCGTCCGCT TACGCACTAG TGGCAAATCG GTTTTTTCAT AAACCGTACG CCACCATTCC	4320
CAGGCAAGCC CGGTACACTC TCTAATTTTG ACAGAGAGAT TACGAACATT CCCTTTTAAA	4380
GGAATACTAG TGGTAAAGTG AGCCGTAAA TCCTGCCCAT TTCTGTCCCA AGCCTTAGGA	4440
GTCAAGACTT CCTTACCTTG ATGATCATAG GATAATTCAT TCCAAGTAAT ATAATATTGG	4500
GCAACATAGG CACCACTATG ATCCAGCAGT AAATCTCCGT TTCTGTAAGC TGTAACCTTA	4560
GTCTCAACAT AGTCTGTAAT ATTTTGAAAG GTCGCAACTA CATTGTACAG TAAAAAGAA	4620
GTTGTATAGG AAATCGGCAA GCCTGGATGA TCTGCTGTAA AGCGACTGCC TTCTTGAATC	4680
AAGTCCTCTA CCATATCCAC CTTGCCTGTT ACAACTCGGG CACCCGAACT TGGGTCGCCC	4740
CCTAAAATAA CCGCCTTCAC TTCTGTATTG TCCAAAATCT GTTTCCACTC TGTCTGAGGA	4800
GCTACCTTGA CTCCTTTTAT CAAAGCTTCA AAAGCAGCCT CTACTTCATC ACTCTTACTC	4860
GTGGTTTCCA ACTTGAGATA GACTTGGCGC CCATAAGCAA CACTCGAAAT ATAGACCAAA	4920
GGACGCTCTG CAGAAATTC TCTCTGTTTT AAATCCTCTA CCGTTACAGT ATCTTGAAC	4980
ACATCTCCTG GATTTTAAAC AGCATCTACG CTGACTGTAT AATAAATCTG CTTAAAATTA	5040
ACAATCTGAA TCTGCTTTTC GCCTGAATGG ACAGAGTTAA AATCAATATC AAGAGAATTC	5100
CCTGTCTTTT CAAAGTCAGA ACCAAACTTG ACCTTGAGTT GTTCCATGCT GTGAGCCGTG	5160
ATTTTTTCAT ACTGCATTCT AGCTGGGACA TTATTGACCT GACCATAATC TTGATGCCAC	5220
TTAGCCAACA AATCGTTTAC CGCTCCGCGA ACACTTGAAT TGCTGGGGTC TTCCACTTGG	5280
AGAAAGCTAT CGCTACTTGC CAAACCAGGC AAATCAATAC TATAAGTCAT CGGAGCACGA	5340
TCGACCGCAA GAAGAGTGGG ATTATTCTCT AACAAGGTCT CATCCACTAC GAGAAGTGCT	5400
CCAGGATAGA GGCGACTGTC GTTGGTAGCT GTTACAGAAA TATCACTTGT ATTTGTGCGAC	5460

293

AAGCTCCGCT TCTTTCTTTC GATAACAACA AACTCAATCGG GTAGCTGATT ACCCTCTTTG	5520
ATGAAACGAT TTTCAATACT TTCTCCCTGA TGGGTCAAGA GTTTCTTTTT ATCGTAATTC	5580
ATAGCTAGTA TAAAGTCATT TACTGCTTTA TTTGCCATCT TCTACCTCCT AATAAGTTCC	5640
TGGATTGAGT TGCATAAACT CAGACTTGTT CAGCGAAATC AGCCGTGGTT GGACTAAGTA	5700
ATCCAAAATT TCCTCGTACA ATTCTTCTGA GACATTGCGT CGCCGTCTGG CTAAATAAGA	5760
AGTCGGAATG ACCGTATTAT CCAACATAAA TACCTTATCT AAGTCAATCA AGGTTGGTCT	5820
TGTAAAAGGA TTACGAGCTA GATCCGGCTC TTCTATCATA AAGTTCTTGA CCAAACGTCT	5880
GGTCAAGAGA GCTGGTTTGA AGGTCTGATT TTAAACCAAC TCTTTGTTTT TAGTCATGCT	5940
GTTGTCAATA CAGATATACA TATGATTCTT CACAGCCAAA TCGCTACTAA TAGTCGGAAA	6000
AGGCCAAATA AGAGCTACAA CATCTCTCTT CTTAATCAAG CAAGAGCACC CCGTTTCTC	6060
CTAATGTAAC ATAGACAGGA TTGACCAAGT CTTCTGATTG ACTCAGAATT TCCAAAGTTT	6120
GAGTTTGGCG CGCTGTCAAT TTAGTAGCAT CTTGTCTCTT CAATACAAAA TGCTTGTCGC	6180
CAATAACCTT GACAATATAA TCCTTCTCCA AAGCTGACTG GTAAATCCAC ATCAGATGTT	6240
GTCTGTCTCG AGAACTCAAG AGAGAAGGAT TTTCAAGCCT CCCGATAGTC TGATAAAAAAT	6300
CAAAAACAGG AGCTAACTCC TGCCAATCTG ATTGGCTAGT TGTCAAGGCT AGAAAAAGGG	6360
CTTTGCGAGC TGATACTTCT TGGTTAGCCT TGAGAGTTAC TTTCCCTCC AAGTTTTTTA	6420
GAAATCGGGA AACTCCAGAA AGCAAATTTT TCTCTAACTG CGAGAAATAA AAACCTTTCC	6480
TTCCCGAGCA TAAGTCTTTC ATGTGCTTTT CTCTAGCAAA TAAGAGCTCA AACATTTGAT	6540
AGTAAAAGAA AAATATCTGG CACTGGGTCG CGCTCATCTT TTCCTTATCG GCTTCTTTTT	6600
TTAACCAGAG CAAGGGCGAC AGGTAGCTGG ATTGAGACAT TTCCTCTACC TCCTACTCTT	6660
TTTAACTGG AGCATCTGCA CTAGCTGCCA CTTCTTTTGA CTGGATACTT TCCCACTGGT	6720
TAATCTCCTC TGAGATAAGA CCTTCGCATG TCTTGACAAA TAGGGCAAAA GCCTTGGTCT	6780
TTCTTGATAA TTTCTCCGTT TGGCATTGAT AGAGGAATTT TTTCTTCTCC AGGAGTTGCG	6840
CAGTTTTTTG GTAAGAAATC CAATTTTCCT TTGCATTATA CAAATTGATA ATCCCTCAC	6900
ACAGCAAGCC GAGACTGGAT AAGGCAACCG AAATCAAACG GTAGCGATCA CCTGGCATAG	6960
GAATAGCACA AAAGACAGCT ATGAGGAAAC CTGCCACGAT TTCTGTTATT TTTAATACCT	7020
TATAGCGCCT ACGATGTTGA ACGCTTTTCT TTAATAAATG AGCTATCTGT ACGTCTAATC	7080
GCTCTGTCAG GTACATTTCT TCTGGCGTCA TATTGTAAC TCCTTTCATT TACTTTGATA	7140
ATCAGGG	7147

294

## (2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 755 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

```

CCGCATGGGA TTGGTGTCTT TTTGGGCAAT CTCTTTGACC AAACCTGGAAA CATGTTTTAT      60
GCGCCTGCCT TTACTGCCCT TGTGGGCGGT ACGTCTATAT GATCCTAGTC GCAAAAAGTTC      120
CCCGCTTTGG AGCCATTACC ACTATCGGCC TTGTCATTGC CCTCTTTTTC TTGGGAACTA      180
AACACGGTGC TGGTTCCTTC CTCCTGGAA TTATCTGTGG CCTCCTAGCA GATGGAGTAG      240
CTCATTTAGG AAAATACAAG GACAAAACAA AGAACTTCCT TTCTTTCATT ATTTTCGCCT      300
TTAGTACAAC AGGACCAATC TTGCTTATGT GGATTGCGCC CAAAGCCTAT ATGGCTACTC      360
TTCTGGCAAG AGGAAAATCC CAAGAATATA TCGACCGTAT CATGGTCGCT CCAAACCTG      420
GAACTGTCCT TCTATTTATC GCAAGTATTG TCATCGGAGC CCTAGTGGGT GCCTTGATTG      480
GACAAGCCTT GAGTAAAAAA TTTGCCCAGA AAATCTGATC AGTAAAAAG AGCCACGCGG      540
CTCTTTTTTA TTTATGGCTC AATTCTTAG TCAAGAAATC TCCCAAGAAT TGGATTGCAA      600
AGATAATCAA AATGATAATA ATGGTTGCCA AGATGGTCAC ATCGTGATTG TAGCGGTTAA      660
ATCCATAAGC GATGGCTACG TTACCGATAC CACCAGCTCC AACCGCACCG GCCATAGCTG      720
TTCCCAACA AGGGAA:CAA GGTcACAGTC GTCAC      755

```

## (2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3010 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```

TTCAATTGGT ATCTCAATCA ACGGTCTTCA CATGGTTTCA ACTGGTTTGA CTCTTGAAAA      60
AGCGAAAGCT GCTGGTTACA ACGCAACTGA AACAGGCTTT AACGATCTTC AAAAACCAGA      120
ATTCATGAAA CATGACAACC ATGAAGTAGC AATTAAGATT GTCTTTGACA AAGATAGCCG      180
TGAAATTCTT GGTGCCCCAA TGGTTTCACA TGATATTGCA ATTAGCATGG GAATCCACAT      240
GTTCTCACTT GCTATCCAAG AGCATGTGAC AATTGATAAA TTGGCATTGA CAGACCTCTT      300

```

295

CTTCTTGCCA CACTTCAACA AACCATACAA CTACATCACA ATGGCTGCCC TTACGGCTGA	360
AAATTAAAAA TGAATGAGCT ATCTGGCCTT AAGTTAAGGT CAGATAGTTT TTAGCTAATT	420
TGTCCCCATA CAATTATAGT TTTTMTATCT TGTGCTTCAT TCTGTTCTGA CTTAAAAATGA	480
AAAGGTAGCT ACCAATACAA ATGATGAGGA TAAAACAAAT GACTGAAAAT CGTTATGAAC	540
TAAATAAAAA CTTGGCACAG ATGCTCAAGG GTGGTGTAT TATGGATGTG CAGAATCCTG	600
AACAGGCTCG TATCGCAGAA GCTGCTGGTG CGGCAGCTGT GATGGCCTTG GAACGAATTC	660
CGGCTGATAT TCGTGCAGCT GGAGGAGTTT CCCGCATGAG CGACCCAAAG ATGATTAAAGG	720
AAATCCAAGA AGCGGTTAGT ATTCCAGTAA TGGCTAAGGT CAGAATCGGG CATTTGTGTG	780
AAGCTCAGAT TTTAGAGGCT ATTGAAATG ATTATATCGA CGAGAGTGAA GTTCTATCTC	840
CAGCTGATGA CCGTTTCCAT GTGGACAAGA AAGAATTCCA AGTTCCTTTT GTCTGTGGTG	900
CTAAGGATTT GGGTGAAGCC TTGCGTCGTA TCGCTGAAGG TGCTTCCATG ATTCGTACCA	960
AAGGAGAACC AGGGACAGGG GATATCGTCC AAGCTGTTCG TCATATGCGT ATGATGAATC	1020
AGGAAATTCTG CCGCATTCAA AACTTACGTG AGGACGAGCT TTATGTTGCT GCCAAGGATT	1080
TGCAAGTCCC TGTAGAATTG GTCCAATATG TTCATGAACA TGGAAAATTG CCAGTTGTAA	1140
ATTTCCGTGC TGGAGGTGTT GCAACGCCAG CAGATGCTGC GTTAATGATG CAATTAGGGG	1200
CAGAGGGGGT CTTTGTCCGT TCAGGTATTT TCAAGTCAGG AGATCCTGTT AAACGAGCGA	1260
GTGCCATTGT TAAGGCTGTG ACTAACTTCC GTAATCCTCA AATCCTAGCT CAAATCTCTG	1320
AAGATTTAGG AGAAGCCATG GTTGGTATTA ATGAAAATGA AATCCAAATT CTCATGGCTG	1380
AACGAGGAAA ATAGATGAAA ATCGGAATAT TGGCCTTGCA AGGGGCCTTT GCAGAACATG	1440
CAAAAGTGCT AGATCAATTA GGTGTCGACA GTGTAGAACT CAGAAATCTA GATGATTTTC	1500
AGCAAGATCA GAGTGACTTG TCGGGTTTGA TTTTGCTGG TGGTGAGTCT ACAACCATGG	1560
GCAAGCTCTT ACGTGACCAG AACATGCTAC TTCCCATCCG AGAAGCCATT CTATCTGGCT	1620
TACCAGTGTT TGGGACCTGT GCGGGCTTAA TTTTGCTGGC TAAGGAAATC ACTTCTCAGA	1680
AAGAGAGTCA TCTAGGAACT ATGGATATGG TGGTCGAGCG TAATGCTTAT GGGCGCCAAT	1740
TAGGAAGTTT CTACACGGAA GCAGAATGTA AGGGAGTTGG CAAGATTCCA ATGACCTTTA	1800
TCCGTGGTCC GATTATCAGT AGTGTTGGTG AGGGGTGAGA AATTTTAGCA ACAGTGAACA	1860
ATCAAATTGT TGCAGCCCAA GAAAAAATA TGTTGGTAAG TTCTTTTCAT CCAGAATTGA	1920
CTGATGATGT GCGCTTGAC CAGTACTTTA TCAATATGTG TAAAGAAAAA AGTTGAGATT	1980
GAATTTCTCA ACTTTTTCAC ATGTAATAAA CAATAGCGAT GTATTGAAGT GCGGACGCAG	2040

296

CTAGGATAAA GAGATGCCAA ATCATGTGGA AATAAGGTTT TTTCTTGGCA TAAAATCCAG	2100
CTCCAACGTG ATAACAGAGT CCGCCAGTTA CCATGAGACT CCAGAAAACG GGTGTCGTTT	2160
GACTGATAAT GGCAGGAATG ATAGCCAGAA CCAACCAGCC CATAATCAGG TAAAGAGCAA	2220
GGCTAAATTT CTCATTGACC TTTTATAGCA AGATTTTATA GAGAATACCA AAGATGGTCG	2280
TTCCCCATTG GATGACAATA ATCAGATAGC CAAACCAGTT ATTCATCAAG GTCAAGACAA	2340
CGGGCGTGTA TGAGCCGGCA ATGGCAACGT AAATCATAGA ATGGTCAATG ATTCGCAAAA	2400
CATATTTGTG GGTGGAACCA TAGGCCATAG AGTGATAAAT GGTGGATGAT AGGAACATGA	2460
GAAAGAGACT GATGACGAAA ATGGAAACGC CGATAGAGGA TAAAAATCCG TGTGCTTCAT	2520
AACTATAGAT GGATGAAATA GGCAGCAAGA TAAGCATGAT GACTGCACCC ACAGCATGGG	2580
TCACGCTATT AGCAATCTCC TCTCCAAAAC TGAGTTGTTT GCTGAGTTTA AGACTAGTGT	2640
TCATTTGATT ACCTCCTCTT GAGTATGATC GATTAAGTCT AGAGTTTGAT GATAGAGTTT	2700
AACGGTTTGG CAGCTGGTTT GGATAATAGG GTTAGCTGGG TCAATTCCTT GGTTTCATGA	2760
GTCCACAAAA GCATCGTAGA GTTGGTCTGA ACTTGCTTGA GTTTGTAGAG TATTAAGTGT	2820
CTGGGCTATT TCTTGAATAG AAAATACAGA CTTGAGGGTT GTGATAGCAA TCAAACGGGC	2880
AATCTGTTGG CGTTGGTATT TTTTTTTGTC AGGCTTTGTC AGGTAACCAT TTTTCACATA	2940
ATTGTTGACC ATAGATGCTG TTAGGCCCTT GTCTTTATTA GGAGAGATAG GGGCGCAGAC	3000
CTGATTGACA	3010

## (2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15213 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CATAAATCGG TGCAAATAAC TTAATAGTGA AGTAGCCATT TCTTTCGTAT TTACCTGAGG	60
CATATTCCTT AGACGAAAGA ATATTATTAT CAATCAAATC ATTGAATGAA CGTAGTCTTT	120
CAACTTCTTC TACTGTTAGA TTTCTGACAA CATTGTGTC ATAGACCTTA TTTCCATCAG	180
GATCAGGATG GTAATCATTT GTAACCTTTC TAAGAAGTTG TTGTTTTTGA TTCGTATCCA	240
ATTTAAGAAT TGAATTTTCT TCGAGATATT CCAACATATA AACAACGTCA AACATGTTGT	300
GGACATATTG CTTCAAATCA TCTGCATTAT TAAATCTTGT AGTTGGATCA AGTACTTGTA	360
ATCGTCGACT TTCTGTACTA TCAGATTTTG AATGTTTCAA GATGGAGTTG ATGGTAATGG	420

297

TCGCATCATC TGGATGGTCT GGTGCTTGTA ATAATCCTTT AGCAAAGAAC TCTGGTCCCA	480
AGCCACTTCT TCGACCATAT CCTCCAAGAT AAATGTCCTG ATCTGAGTCA TGTGTCATCT	540
CATGCGTATA AGTAATAGCT CCATCCTTAT CCAACATTCC ATAACCCATA TAATAAACTG	600
CATCACCTGT AGCATAAGCA CCGTGTTGAT TATGCCCAAC TTTATTTCCA ACAGGTCCAA	660
AGAAATGTTG CATTGCAGGA TTTGGATTAT CAAAATCTGC CACTTCTGTA GCTTTCCCTA	720
CGGTATTATC ATCGCCAAAT TTATAAGCAT CGTAAAGCAA AATATTTCTA TAAAGTTTTT	780
CACGTGCATT GTCGTCTAAA ATACGATACC AATAATCGTA GTGATCTCGC TGACGTTTGG	840
CTGTTTCACG CGCATTTTCT TCAACAAAAT CATTGAGAGC CTTGCCCGCT TTATGGTCAC	900
TACTGCGGTA GCGATCATAA GCTCCAAATC CTAGACTAGA CATGGTCGAG ATGACAAATA	960
CGGATCTCTC TGGCAAGGTC AGGAGAGGCA AGACCATATT GCGGTATTTC CATGTGGCAC	1020
TCGTGATACG ATCATAAACA CCGATAGAAT ACTTGGTGCC AGCTAACCCT TGCTTCGTTT	1080
TCACCTCTTC GATAGTGGAT TTTTCTTCGA CAATGTAAGC CTTAGTCTCT GATTTAACC	1140
AGTCATTATT GCTTGTATTT GGTAAAAAGA CTTTTCGGTA ATGTTCCAGC GTGCTAAACA	1200
AATCTGTCGT TCCATGTGTA CTGGCAAGAC TGATACCATA AGTATCGACA TTATTCTTAG	1260
CTAGAAGATT GTTAAAGCCA GATTTACCCA ACTCAATCAG AGTATCTAAT GGTGAAGCAT	1320
TCCCCTTACC AAAGAAGTCC AAATGGTACA GAACTAGGTC TTTGACATTC ACCTGACCAT	1380
AGCTAAAGTT ATACCACCGT TCCAGATAGG TCAAGCCAAG TAGCAAGGCT TCCTTGTGTC	1440
GTGTGATTTT ATCTACAAGA TAACCTTCAG TGACGGGGTT AGCACTAGCC AGTCCAGCAT	1500
CCGCTGACAA GAGTTTTTTC AAACTGTCTT CCAGTTGTTG TTTTGTTTTG GCGAACTGGT	1560
CTTCTAGATA GAGCTCAGTT TGCTTGACGT TTGGAGAAAT ACCCAGCGTC TTTCTGATGG	1620
CTTCTGAATG ATAGTCAACC TTTTGTAACT CAGGTAAGAC TTGCTTGATG ATAGAGGTTT	1680
GGTCATACAG GAATTGGTTT GGCGTATAGA GAAGTCCAGT ATTGCCCAGA CTATATTCTG	1740
CTAATTTGGC GAAATCATTC TGGTATTGTA GATCCAGCTT CTCAGATAAA TCATCCTTGT	1800
AGTGAAGCAA GAGTTTGT TTGCACTCTGTT TGTTAGAAAC AATGTCTGTG ATGACTTGGT	1860
TGTCCTTCAT CATGACTGCT GACAAGAGTT CTTTTTGATA TAAAAGACTG TTCTCATTTGA	1920
CCAGGTTTCC GTATTTGACG ATGGTTGCCT TGTTGTAGAA AGGTAGCAAT TTTTCAATGT	1980
TTTTATAAGT CAAGTTGCGC TTAGCTTGAT AATAGGCCAC CTTAGAAAAA TCACTGTCTT	2040
TTTTGCCACT TGTGAAAGT GGCTCCACTG TTGGTAAAAT GAGAGGATTG ATTTCTGCTT	2100
TTTTGCTTGC AATTTGAGAA GCATCTAGCA TTGTTCTCTT TTCTTCAAAG GATTCTTGC	2160



298

TGACGACCTC ATCCTTGACC AAGGTGACAT TGTAGACTCT GTTGGCCTTG CTGCTGAATG	2220
TGTCCTTTAC CTTCATTTTCG TTATAGTGGT AACCAGTGAT GGCATTTCCG TTGGTTACAT	2280
TAACATCGCT GAGAACATTG GTCAAACCTC CAGCATGCCT AACATCACCA GAAGTTTCGAT	2340
CCCACAAATT GCCTGCCACT CCAGCGACTC TACCAAAGTG CTTGACATTG TTGATATCAC	2400
CTTCAGCATA GCTATCTTGG ATCTGTGCAT CTCGGTCTAC TAGGCCTGCA AGTCCACCCA	2460
CAGTCTGATC TGAAGTATTT GTGTTAGATG AAATGGCTAC TGTGCTTTT GACTTAGTAA	2520
GTAAAGCCTT GTCACCTGTC AAATGACCGA CCATACCACC GATATTGTAG GCAGCAGTCG	2580
TTTCATAAGT GTTGATAATT CTTCCTTGA AACTGCTCTC TGTGATGCTT GATTGCTCAG	2640
CCTTAGCCAG CAAACCACCG ATACCACGTT CACCAGCCAG AACACCATCG ACGTGAACCTT	2700
GCTTAATTTT TGTGTTATTC TGAGCTTCAT TTGCCAGTGA ACCGATATCA TCTTTCCCTG	2760
AAATAGCAAC ATTTTTTAGA CTCAGTTTTT CTA CTGTAGC ACCACTCAAG TTTTCAAACA	2820
GAGGTTTTTT CAAATTATAG ATAGCATAAT TCTTGCCATC TTTTTCACCG ATTAAACGAC	2880
CAGTAAAGGT GTCCTTGATA TAGGATCTTT CATCAGGACC AAGCTCCACT TCGTTAGCAT	2940
TCAGGCTGGC CGCTAAATGA TAGGTTCCAG AGGGATTTTG GTTTATAGCT TTGACCAGAT	3000
TACTAAAGGA AGTAAAGTTT GTTGTTCCTT CTGTTCCCTT CTTAGCTAGA TAGAAGGTAA	3060
AATTATCTTT ATATCTGCTT TCTATCTCCT GCTGAAGCTT CTCTACTTTT GCTGTGATTT	3120
TATAAAGGAT TTTATCATTT TTTCTTTCCT CTGATATTGA TGCTACTGGT AGGTATACAT	3180
CTTTGAATGA AGAAGATTTT ACTTTAACAA AGTAGCTATT TGGATTGCTT GGAACCTTGCT	3240
CTAACGAAAT GTGTGTTTA TAAGTACCAT TTGACAACT GTATAACTCT AGGTCGGAAA	3300
CATTTCTTAA TTCAAGTGTT TTCTCTGGTT CTCTACCTT TTTATCAGGG TCTAGTTTAT	3360
TTTCTTGTTT AATTTCTTCG TTCCATTG AATTGGATGT GTTTGATTGG GTTGAAACAT	3420
CCTCAGTTGA ATTTCCGTTT GATGGTTCTG GTTCTGTTTG TCCATTCTCT GATGTTGTAT	3480
TACCTGAATT TTCTGGTTTT GTTGCAGTTC CGTTTTTTC TGGTTGATTT GATTCTTCAA	3540
CTGGTGGTTT TGAATCACTA GGTATTATTG ATACTTCTCC AGTATTTTCG TTAGCTATTT	3600
TCCCAGAGTT TGTGTTGTTT TCTTCTGCAG GTTGAACCTG TTTTCTGTT TCTTGATTTG	3660
AGGTACCTTC TACTGTGCCT TCATTTGGAT TTA CTGGAAC TTCTTCTACA GTTTTTTCTG	3720
AATTTTCATT TTTAGAGTCA TTATGTTCTG GTTTATTGTA TTCTCCAAC TGGGTTGTCG	3780
AATCACTAGG ATTA CTGGAC ACTTCCCCAG TATTTTGCT AGATGTATCT GGTGATACTT	3840
TCTCTGAATT CGTTGTTGAT TCTTCTGCAG GTTGAACCTG ATTTTCTGCT TCTTGAATTG	3900
AGGTTCCCTC TGTAGTACCT TCATTTGGAT TTA CTGGTGT TTCTTCTGTT GGTTTTACTG	3960

GAACCTCTTC AGTTTTTCTT GGACCTTGTT CTTTGGTCTT CTCAACCGGA GTTTCAGGTT	4020
TTACTTGCTC AATATTACCC TTATATTCTG GAAGCGGTGC TACCTGCTCT GGTTACACCTT	4080
TATCACTTAC CACAGTATCT GCGGACTCTG GTTGAACCTC AGTCTCACCT TTGTCCGTCA	4140
CAACTGCTTC GGGTAATGTA GGTGAACTT CTGGTTCGCC TTTGTCACCT ACTACAGCTT	4200
CGGGCAACTC AGGCTGAATT GCGGGTTCAG CAATAGCTCC AGACTGTACG TCCTTATGTT	4260
CTACACCAGT CTCAGGTGTG TCCTTTATAA CTTGAGTTTT TTTAGTACCT TTTTCGACTA	4320
TTCTTGGACT AGGCGCAGTC GTTGAAGTTG AAACAATTTT TCGCGAAACT TCTTCCTGT	4380
TTACAGAGAA TATTCTGACG ATTTCAACTT TCTTACCTAA TTTACCTTCT TGTMTTACTC	4440
TTACAGTTCC TTCAGTAAA TCAGGATTTT CTTGAATTTT TTCTTGAAAA TCTATTTTTG	4500
TCTCCATAGT TTCCTCACGA TATAAGAGTT CAGGTTTGTT CAATTGACCT GATAAAACTT	4560
CATCCTGTGG ATTTAATGTA TTTACCCAG TCTTTTCTTT TGGAGAAATC TTCTCCTCTT	4620
TCTTCGTTTC TAGATTCTTA TGTTCCGCTA ATTGTTCTTG AGAATCTGAA GATTGTTTCT	4680
CTTCTTTTCT TGGATTGATT AATTCAGTAG AGAAAGGTTT TTCAACTACT TGAACCTCTG	4740
TCGGCTTAGT TGAAGAAACA GGTGTTTGTT CCTGAATAGC TTGTACTGTT GATGGATGGT	4800
CTACAAAATT CCGTGTAACA TTATAATCCA CCTTTTGTTG TTTGTAGGA GTGGCAACTG	4860
AACTCTTTTG ATTACTTACT TCAGACTCAG AAGTCGTTTT TCCCTCTTTG ATATATCCAA	4920
TATAAGTGTA ACCTGAAATC TCTTTAGGAA GAGGTAATTT TTCTCCAGAG GTCAATTCAT	4980
AGTCCGTATT GTAATTTAGC AAAAGATGAT TTTCTAAAGC ATGGACTGAA ACTAAGACAC	5040
CATTTCTAT CCCTGCAACC AATACTAAAT GTAATACCGT TTTATTCTTA ACCTTTTCTT	5100
TGGAAACAGC AAAAATTAAA ATTCCCATAG CAGCTAAGCT AGCACCAGCA AATAGGGCTT	5160
GCCTCTCATT CTTGCTTCCA GTATTTGGCA ATTCCGCCAG TTGATTTTGA GAATTTAACT	5220
TATAAACAAG ATAATAAGTT TCATCATCAT TCTCCACGTA TGTCCGAATA TCATAGACAA	5280
GCTGCTTCTT TTCTTCTGAT GATAGCTCTG AATCTGCCAC ATATTTATAG TGAACCTCCG	5340
CAGTTTCTTG AGCATCCACA GATGAACTAG CTAATACAGA CATAAAAAAT AAACCTGAAA	5400
TCGTTGCAGA TACAAGTCCT ACTGATAATT TTCTAAATGA AAAACGCTCT TGTTTTTCAC	5460
CAAAATACTT TTCCATTATT CCTCCTTGAA ATAAAATTTA TATATGTTAC AAAGACCTTT	5520
ATTATATTAG TGTATTATCT ATTATCTATA GAAAAGGCAG TATACCTTAA TTATACTCTT	5580
AATTTACAAA AAAGTCTTAA AATTGAGATG CGCTTTCATA CTTTGTTTTA TATTATTGG	5640
AGGTACAATA ACACCTACCA TGAAATTTAC ACGGTAGGTG TTACTCATAT CACTAATCGT	5700

300

TCTAAAAATG GTTTGAGGCA GTTGAGGAGA ATTCCCTCTA TCCAGCTTCC TTGTGCTGAT	5760
GAGCGATGGT CTTCTGCAG GCTTTTTTTT AGAAAATCTC GGACTTGTTT TGGTGGGATT	5820
TCAAATTCAA AGGCTTTCAT TTTATAGAAA AAGTCGATGA GATGATCTGA CAGGTATTCA	5880
GTTGAAAAGG GTACTTCACC ACTTTTTCTA TATTCTAATA AGAGTCTAGA AAATCGAGCT	5940
TTTTCTTCAG GAAGCTCAG AAAATAGGAA TTGAGGATCC AAGTCTGCTT CTGTTTCTT	6000
TCAATGGAT CCTGACTGGC AATTCGTTGG TCTTTTTCCA GCTCTTTTG GTATTGTTTG	6060
GCCTTGATAG CTCGTTCTGC TCTATTTTTC CAAAAAGAA TTTTTTCCCA CTTGCGTTCT	6120
TCTTGAGTCA GGGTCTCTGT AAAGCCAAAG TAATCTTGAT AAGCACGCTC TCGGGGTCCC	6180
ATGGCTAGAA CCAGATTGTC TGCATATTGC TTGGCGATTT TATCCCTCTT CTTGCGTTCT	6240
TTCTCTGCCT GGATACGGAG TTCTTGTTCC TAGTCAATTT TCTCCTGCC TAGCTTGACA	6300
AGGTAGAGTT GGTCAATCCA TTTCCCAAGT AAAAAGGGTT TGATACACTT TTCAAGGACT	6360
TCTTCCATCC GAGCCTTTTT CTTTGGTTCC GCCTTGGTCC AACTTCCTCC CTGAAAGACT	6420
TCTAGGAAAA GCTGGTAGTC TCTCTCAGGC GCAAATTGAT TGCCACGATT GGGTTTGAAA	6480
ACACCTTTTT CCCAGAGCCA TTTTAGAAGT CGCTCGTCAA AGTTACTTTT ATTGACCTTG	6540
ATTTTTTCCT TTTTCTGAGC TTTTCTGGTT AGATTTTCAA CCTTCTGAG CAGTTTTTCT	6600
TCCTCTTCCA ATTGCTGGTC AAGGGACAAT CGATGAAAAT GACGAACACA GTCGCTACCA	6660
ATTGGAAGA GCGGTTGGCC TGTGACACCG TTAAGAGTT CATAAGCGTA TTTGATGGCA	6720
TTTCCACAGA CACAATTGCT ACGGCCGATA CCGTAAAAA TAAAGGAAAC TTCATTCCAT	6780
TCCTTGGTAG CTTGTTCCA AGTATCCGCT TTCGAAGCCT GTAAACTGC ATCGTGACGG	6840
GATTTTCTAA CTGGAAGTGT CATGAGGTCT CTTTCTAAT ACTCAATAAA AATCAAAGAG	6900
CAAAC TAGAA AGCTAGCCGC AATCAGCTCA AAACACTGTT TTGAGGTTGT AGATAGAACT	6960
GACGAAGTCA GCTCAAAACA CTGTTTGTAG GTTGTTGATA GAACTGACGA AGTCAGTAAC	7020
CATATATACA GCAAGGCGAA GCTGACGTGG TTTGAAGAGA TTTCAAAGA GTATAAGTTA	7080
TACTTTTACA ACTTGAACCT CGTCTTTACC GAGTAAATC AAGTATTTT CAATATTTTC	7140
AATCGAATAG GCTCGTGATA AAGCCTCTTC GTATAGAGCT AACTGACCAC GATAGCGGTC	7200
TACGAGTTGA CTTGGTTTAT CATAGCGGTC TGTCTGTAG TCGAACAGAA CAATTTTGT	7260
TTCTGTAAGC AGATAGCCAT CAAGGATACC ACGGACAACA AAGTCTTCCT GACTCTTTTG	7320
GTCTCGTTTG AGCATGGAGA AAGGTTGCTC CGGATAAAGA TGGTCGGTAT TAGCAAGAAT	7380
TTCTGACCG AGTACTGTGT CAAAGAAAGC AAGAATTTT TCAAGATTGA TCTGTCTCT	7440
GACAGCTTGG CTAGTTTGAA CTTGTTTGAG TGTCTCTGT AGGCTAGCAA GGGTTAGTTG	7500

301

CTGGCTGAGG TCAATTCTCT GCATGAGTTC GTGAGTAGCA CTACCAATCT CAGCTCCAGT	7560
TACCTTTTCT TTGGTTGAAA AATCTGGCAA ATCGAAGCTG ATTTTCTTGC CTA CTGACTG	7520
ACCTTGACCA GCAATCTCGA CACCTTCCAT ATCCATAACT GGTTCTGTAGA ATTTCTTGAT	7680
TTGACTTGGG GTTGAACAC TAGGAAGTTC AATAGCTGCG CCGTGAAGAG TATTATAAAC	7740
TTCCACCTCC TTCAGCATT TCCAGAGCTTC TTGATGGTA TCTGACTGAC GATTGTCTGC	7800
TTGGGAGCTA TCTTGGAGAG GACTCTTGGT TTCCAACCTCT CCGATAGCTT CTCTGGTCAA	7860
CTGATCTTCG CCAATAAAAC GATAACTAAA GTTGAGCTTG TCCTTAGTAA ACACCTTTACT	7920
GATAGCCCAA AGCCAATCTT GGAAATTCG TGCTTGCACT CTAGTATTGC TATTTAGTTT	7980
CCCATTTTTG GCTGCTGGGT ATTCCTTGGG TTCCAGCTTT TCACGAGAAC CCTTGCCGAC	8040
AAGATAGAGC TTTTCTCAG CCCGCGTCAT AGCAACATAC AGCAAACGCA TCTGCTCAGA	8100
ATAGCTTGCT AGCTGTAATT CCTCTTCGTT CTGCTATAG GTCAGACTAG GAATGGAGAG	8160
TTTGATGGTT TTAGGATAGT GGTCTTCTAC TGCCCTGTG TCCATCTTGG CAATATATTT	8220
GACACCAAGA CCATTCTGAC GACTGAGAAT GACTTCTGAC ATAGAGTCTT GCTTGTGAA	8280
ATCTTGATCC ATATTGAGGA TAAAGACGTA AGGAAACTCC AGCCCTTTAC TCTTGTGGAT	8340
GGTCATGAGC TCTACTGCAT CTTTGGCGG TGCGACGGCC ACGCTTGCCA AATCGTGCTG	8400
GGCTTCTAAG ACTTGGTCAA TCATACGAAT AAAACGCGAC AAACCTTTGA AATTGCTCTT	8460
TTCAAATTGA TCAGCACGCA GTGCTAGGGC ATAGAGATTG GCCTGCCTAG CAGGACCATT	8520
CGGCAAAGCC CCAACATAGT CATAATAAAA ACGGCTCTG TAAATCTTCC AAATCAAGTC	8580
ATAGAGAGAG TGGGTTTTGG CATAAAGCG CCAAGAAGCT AGGATATCCA TGAATTGCTT	8640
TAGTTTTTCA GCTAGAGCTG TGTGAATCAA GCCTTTTTGA CTA CTTGCCA TTTTTGTGC	8700
ATTGACCAGT TTCTCATAGA GATTTTCTG GATTTTATCT TCTGCTTTCT GAAGGGACAA	8760
ACGTGCTAGC TCATCCTCAT CAAAACCAAA CATTGGAGAC TTCATAAGGG CAACCAAGGC	8820
GTAGTCTTGC AGGGGATTGT GAATGACACG AAGAGTGTCT AGCATGACTT GCACCTCTAG	8880
GGATTGGAGA TAATTGTTTT GCTCTCCGTC AGTTTTGACA GGAATCCGT ACTCAGACAG	8940
GGCGAGGAGA ATCTGGTCAT TACGACTGCC GCTGGAGGTC AGAAGGGCAA TTTCTTAAA	9000
GGCAACACCT TTTTCTTGAT GAAGTTTCAG AATCTCCTTG ATA ACTAAGC GCATTTGCCC	9060
TGTTAGTTTC GTTTCTGTTT GACTCTCTTC TTCCTCACCT GTATCGTCTT TGTCGTAGAG	9120
GAGAAATGCT GCCTTGTGTT CTGGATTGGG AGTCAGTTTG GTATTGGCAA AAACAAGCTG	9180
GTGCTTGTTA TCATAGTTGA TTTGCGCGAC CTCTTGCTCC ATGAGACGTT CAAAGACATC	9240

302

ATTGGTTGCT GACAGCACTT CTGAACTACT ACGGAAATTT TCCTTGAGGA TAATGAGCCT	9300
GCCTTCTTGG GGATTTTGCG CATAGCGTTG GAATTTCTCA TTGAAAATCT GCGGGTCTGC	9360
CTGACGGAAA CGATAGATGG ATTGCTTGAT ATCTCCCACC ATAAAGCGAT TGTGGCCATT	9420
AGACAACAAT TCCAGCATCC GTTCTTGAAT ATGTTTGGTA TCCTGATACT CATCGACCAT	9480
GACTTCATGG AAGCGCTCCT GATAAGACTC ACGAACTTGT GGGAAATCT CTAAAATCTC	9540
AATGGTGTA TGGCTGATAT CAGCGAATTC GAAGGCATTT TCCTGTCTGT TTCTCTGACG	9600
ATAAGCCTCT ACAAATCGC TCATGAAAGA TTGGAAGGTT TTAGCTAGTT TCCAAGTGTC	9660
TCCATGATAA CGTCTTGAT AGTCGAGAAT CGCTATCTGG TCTGATAATT GTCCTAGTTT	9720
AGCAAATCTG GTCTTTCTCT CTTCGTTGTA GGCATCAGCC AGGGGCTTCA AATCAGCCTA	9780
CGGCTGGCAT TAGTCAGAGC TCGACCGTTT TTCTCCTTAG AGATGGCGAC AACACGCGCA	9840
AGCACTGCCT GATAAGCCTG ACTATCGGAC TCCTGATTTA GGGAGCCAAT TTCATCCAGA	9900
ATTAAGTGAA CATTTTCTAA ATAGGCAGCC TTTGCAAAT CCTTGGCATC GTTATCCAGA	9960
TGGTAACGGA AAAAGCTTTC CAAATCCCAA AGGGCTTGTG TGATTTGCTC GGTCACTTTT	10020
TCTTTTTCAC TGGTAAATC AGCTTTCTCA AATCCTTTGA GGAAAGATTC ACTCAGCCAC	10080
TTTTGAGGAT TACTGGTGGA TTGGAGGAAG TCATAGATTT TATAGACCTG CTGGCGCAGA	10140
CCCCGTTCGT CCTTGCCAGC CCCAGCAAAG TTTTTCAGCA AATGACTAAA GGTCTCTTTC	10200
TGTTTACCTT GGTAATGCGC TTCAAAGACC TCATGAAAGA CTTCGTTTTC GAGAATAAGT	10260
TGCTCGCTTT GGTTTTGTA AATACGGAAA TTAGGTGCAA TATCAAGCAG ATAACCATGT	10320
TTGCCAAGGA ATTTTGTCT GAAAGAATCC ATGGTTCCAA TGGCAGCGTT GGGTAGGTCT	10380
GCCAACTGGC GACCCAAGTG TTGTTTGAGG TCGACATCAT CTGTTTCTTG GATTTTCTTG	10440
CTGATTTTTT TCTCTAAACG TTCTTTAAGT TCAGTTGCAG CCTTGACGGT AAAGGTTGAG	10500
ATAAAGAGTT GAGAAATTC GACACCACGC GCCAATTGGT CCAGAATGCG CTCTGCCATG	10560
ACAAAGGTCT TTCCAGAACC AGCCGATGCT GAGACCAGGA TATTCTGGGC AGAAGTGTAG	10620
ATAGCTTCGA TTTGCTCGGC AGTTTCTTTC TGTTCTTGC TCGAAATTGC TTCTGCTTCT	10680
TGCAGTTTTT GAATCTCCTC CTCACTTAAA AAGGGAATAA CCTTCATCGA TTCAACTCCT	10740
CTCTTATTTT TTCAAGCCAA GCTTGCTTGA GTTTTCTCC GACCAGACGC TTGCCATCAG	10800
CTAGGTCCAA CTTTTCTAGG AAACGGGCTT GGCCAGATG GTAATTGGCT TCAAAGCCTG	10860
TAATAGCCTG ATGTTGCTGG ACGTATGGGG CAATGCTTCT GCCATTTTCA GTATAAGGAT	10920
TGATGGCGAA CCGGCCTGCT AAAATCTTCT CAGCAGCTTT CTTGTAAAGA TAGGCATTGT	10980
AGTCCAGTAG GAGCTGAAAT TCCTCATCTG TCAGTTGATT AGCCTTGTTT TTGTTATAAA	11040

303

ATTTCGCCTAA ATAAGTCTT TCTTTTCCCA AGAAGAGCCC TTGGTATTTC ATAGATTTCG	11100
TGGCTTCTAC CACTGCTCCT GCCAGACTTT TTACCGCCAT CAGAGATTGG ACAGGTTTCAG	11160
CCATTTCCTAA GTACATGGCG CCGAAAAAGT TCTGCTCCCC TTCTCTTTT AGGGCAGCAA	11220
GATAGGTTGG TAACTGAGAA TTGAGCCCAT TAAAGAAATG AGGAACTGG AACTGAGTCA	11280
GACTGGATTT GTAGTCTACT ACTCCTATCG CTCCATTAGC TTCAAACGG TCAATCCGGT	11340
CCACCTTGCC TCGTACAAAG AACTGCGTC CATTGTCTAA TTGAATAAAG GCTTGGTCTT	11400
TTCCACCAAA ATTTGCTTCT TCTTTGATGG TTTCGATGGC TGGATTGTGT CGGAGAATAT	11460
GTCCAGTTGT CCGTGCAACA TCAAGCAAAA CTTCCTTGGT AAAGTGGGT TCCAACTTT	11520
CTTGATAAAT AGCTTCAAAT TCGCGTTCTT GACTGGTTTC TTGAATAGCT TGTTCTAGAC	11580
GTTGGTCAAA GGAATCTTCA TTAGGCAACT GTAAGGCGCG TTCAAAGATA CGATGCAAGA	11640
AATCCCGTG ACTACGGGCA TCAGGATGCA AACGTAATTC CTCCTGCAAG CCTAAAACGT	11700
AGCGTAGGAA ATAAGTGTAT TCATTGCGAT AAAACTCTGT CAAACCCGAC GTAGACAGGT	11760
AAAACCTCTG TTTGGCAGGA TAGAGAGCTT GCAAGGTGTC CTTGGCTAAG GTCTTGCTGC	11820
TTGGACTGGT TGGGATAGCT GGATTTTCCA GACCTTGCTG ATCTAGTTTT TTACCTATGA	11880
CACGCGACAG AACCTTGACA AAAGTCAAAT CTGTGCTAGT ATCGCTCATC TCACCCTGCT	11940
GGTGATAGGC AACCAGACTA GACAAAAGAC TGTGATAGGA CCCCATATCC TCCTTAGACA	12000
GTCTTTTGTG ATTCATCTCT TCTCTCTCC GCCTAAATCC AAAATGGATC AACTCTTGAA	12060
GATAGGCAGA TTCCTTACTT TCACCTTCGT TAAAAAGGCT TGGAGCCGAC AAGAACAAC	12120
GCTTACGAGC AGAATTGACC AAGGAAAGCA TAGGTAGCG ATTTTCTTG AGATTTTCAC	12180
TCTGGCAAT CAGTAATTGA ACGCTTCTT CGGTGGCTG GTTCTCTT TCGCTTCTT	12240
CATCTGTCAG AAGACTGGTG TTTTGAGAAA TTTTGGTAA ATTGTCCTGA GTTAGTCCAA	12300
TAGCATAGAC AAAGTCAGCA GTCAATGGTG CAATCAAATC GTAACCTGTC ACCAGAACAG	12360
TGTCCACTGT TGCTGGAATG GTACGGTATT GGGACAAACT CATTCCAGAA TGGAGCAAGG	12420
CTAGGAAGTC TTCCAGACTA ACCTGTGAAC CAGCAAAAAC AGTCGCAAAT TGTCTAAAA	12480
CATGGCAGAA AGCCTTCCAA ACTTCGGCTT GTCTTCTCTG TTCTACAGCT TCCAAAGTGG	12540
TTGTCAAATC TTGTAAGTGC TTGGTCACAG CTCCTCTTT TAGAAAGACA CTCCATTTTT	12600
GTAGGAGTTT TTCAGCCTTT TGTTCGGGC TGGCAAAGAG GGTTCAGA GGTGCTAAAA	12660
TTCTCAGGCG GAGGACATTC AAACGCTCAA GATTAAATTT TCCATGGTGG GATTGGTGA	12720
AGGTTTGCTG AAAGGCTGGC AAGCCATTGA TACCAAGATA GCGGATATAT TGCTCAAAAG	12780

304

CATCAATATC AGACTGACTG AGGTCAGTAT ACAAATCAGT TCTAAGAGAGA TTAATCAAAT	12840
CCTCCTGACG AAAACGGTAA CGTTTAAAG CTAATAGAGA CTCGACAAAC TGAGTCAAGG	12900
GATGATGAGC CATGGCTTCG CTTCTACCAA GATAAAAAGG AATCTGATAC TGGTCAAAAA	12960
TGGTTTTGAG AGATAACTGG TAAGAAGCTA CATCCCCCAA GAGAATACGA AAATGCTTGT	13020
AGCTCAGGTC TGAGTTCTCA TGTAATTTCT GACGAATACT ACGGGCTACT AGCTCCAACT	13080
CCTCCTTTTG CGTCAAACAA GACCAGATTT GTAAATTTTC ACGGTCTTTC TCATCGACAT	13140
CCAAAGCGAG TTCTGAAAAG TCATAAGAAG ACTCCAACAA ACGAGAGGCC TTGTCAAAAC	13200
TATCCATCTT CTCATGAGTT TGAGAACAGT CCTGAGCAGG CGTTTGGTAT TTAGAAGCCA	13260
GATGATGGAG AAATTTTACG CTGGCTTGGT AGAGATTGCC CTCGCTAAAA GGAAGCTGAT	13320
AGGCTTTCTT ACTAGCATAA GCCCGGATAA CAATCTCAAC ACCTTTGCCG TGAAGTAAGT	13380
CCACAACCCG CTCTTCTCA GCAGAAAAAC GAGTAAAGCC GTCAATGACC AAGGCGATT	13440
GATTAAATC ACTACTTACC TTGTCATTCT CAATAGCCTC AATCAAATGG GACAACTGAC	13500
TTTCCTGGGC TAACTGACCT TGATTAAGAT AGGCTGTTAC TTTCTCAAAA ATCAAGAGTA	13560
AATCCGCCCT CTTATCTCA TCTGTTAAAT TCTCCAAGTC CAAAAAATC ATCTGAGATT	13620
TGGTCATCTC ATGGTAAAGC TCAATTAAGT GCTGGATCAA TTGAGGATCC TGCTTAATAG	13680
CGCCATAAAC ACGCAAGTCC TTGGGATCGA GTTCGGCAAG GCATTTGTAA AAGGCCAACC	13740
CAAGACCGAT ATCATCAAGA GTAGTTTATG CTGGTAAATC ATTCAAGACC AGATAGCGAG	13800
CCATTTGAGC AAAGCGCGTG ACGGTAATCG AAAAGAAGC CTGCTGGGAC AAGTATTCCA	13860
GCACGGCGCG TTCCTTTTCA AAAGAAAGAG AGTTGGGGGC AATGTAGAAG ACCCGCTTGC	13920
CAGCTGCAAC TAGCTCTTCT GCCTCTCTTG TTAGAATTTC TGTCAAAGAA GTCCGAATAT	13980
CAGTATAAAG TAATTTATC TCAGCCTCGT TGGAAATTTT CATCACCTTA TATTATACCA	14040
TGATTAGCCT CGTAAATCTG TTAATAATTT TAGGCCATCC TTTCTTTTCT TCATCATCTG	14100
CTAAATCTTA AATACTTAGC TTTACTTGTA TTAGATAGAA TAAGTCTGGC TACTGAAAAT	14160
CACATAATAA AAAAGCCTCG GTAACAAGGC TTTGAGTTTT ATGATTGTTT CTTAGGTACG	14220
GAATACACTT CAATGTGTTG TCCCAGTATC TTAATGTCGA CTGGTAGATT GTCTGATTTA	14280
TGCCCATCAA CATCGGACTC TAATTCGATA TCAGAAGAAG TTTTAATATT ACGTGCCTTT	14340
ATATATTCAA TATTCTTGAT AGAATGATTG AACTATAGTA AATTGAAACT ATAATAGTAC	14400
ACCGTGGATG CTAATAATTT TCTAGAAATT AATTTGATTT CCCTAATCAA GCTATTCTGA	14460
TCTTATTTCA ATCTACTATA ATAAAATGAA CCAAAAATAG TACACAATGT GGTATAATCT	14520
TCTTATGGCA TATTCAATAG ATTTTCGTAA AAAAGTTCTC TCTTATTGTG AGCGAACAGG	14580

305

TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATT ATGGCTGGTT	14640
AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAACAA AACCAAGAAA	14700
AGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGATGCTT ATTTGACTGA	14760
AATAGCTTCT GACTTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG	14820
CTACACTCGA AAAAAAGAAC CACACCTACT ATGAACAAGA CCCAGAAAAA GTAGCCTTAT	14880
TTCTTAAGAA TTTTAATAGT TTAAAGCACC TAGCACCTGT TTAGATTGAC GAAACAGGAT	14940
TCGATACTTA TTTTATCGA GAATATGGTC GCTCATTAAG AGGTCAGTTA ATAAGAGGCA	15000
AAGTATCTGG AAGAAGATAT CAGAGGATTT CTTTGGTTGC AGGTCTAACA AATGGTGAAT	15060
TAATCGCTCC AATGACTTAC GAAGAGACGA TGACGAGCGA CTTTTTTGAA GCTTGGTTTC	15120
AGAAGTTTCT CTTACCAACA TTAACCACAC CATCGGTTAT TATAGTAAAA TGAAATAAGA	15180
ATAGGGGGGG GGGGGGAGGG GGGGGGAGGG AGA	15213

(2) INFORMATION FOR SEQ ID NO: 27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6004 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

TTATTACCTG AAACATTAAA TTTAATTGGA CATCCCGTTA TCAATTTTAT AATATCATCA	50
AGATTTTTAT TATCTGATTC AGGAATTTTA TCTGATATAA CAACACCATT TTCAAGATAG	120
TTCAATTAAT TATTTGATTC ACTAACATTA GTGTTTTGAT CTCTATCAGG CCAAAAATAA	190
TGGTTATCGG AATCTAAATA CGATGAGTTT AAAATATTAT TACAAATTAT TTGATTTGCT	240
CCACCAGGAA TATATCTCAC TACTAAATTC TGTTTAAGAT TCTCACTACC TGAATGAGTG	300
ATAACAAACT CTAGAATATA TTTAGCTAGT CTATCTTCAA CATAAATCAT CTTCTAGAA	360
TGATACACAT CACCTAATTC AAAAAATGCA TCCTGATAAT CAATATTTTC AATAACATCT	420
ACCTTTTCTC CGTTTTTCAC TAAAAGTTTC ACGGCTTCTC TAGGAAAATC TTTTATAAGT	480
TGTGTAGAAT GTGTAGTGAT AATAATTTGA TGTTTTTTAT TTAAACACTC TTGAAGTAAA	540
AACTCTTTAA ATTTATAGAT TGCACTCGGA TGAAGTGAGA TTTCAGGTTT ATCTATTAAT	600
ATTAATGAAT TTGATTGCGC ATTTACTATA TCATTTACTA ACAAATAAT TCTAGCCTCA	660
CCTGTTCTTG CAAAAGCCTC GGAATATCTT TTTCCAGATT TTTTCATCCA AATAGTTTGT	720



306

GAAGCTTTTA	TATCATCACC	TTTTGAATAC	AACCTATGTG	TTAAAAATTG	AATGTCTGTA	730
TAAGATTCAT	CCATTATTTT	ACTAATAATT	TCACAAACTT	TATCATCAAC	TTTAACATTA	840
TCTATAACCA	TTTCCTTTTT	ATAACGCGTA	TAGCTACTTG	TATTATTCTT	TAAAAATATCA	900
GCAACTGGCT	TAGATCGTAA	TCTTATAAAA	TCTTGTTTAC	TACGTTGAGT	AGAAATTTTT	960
TTAAAAATTAT	AGTGATAGAA	AAATAAATCA	AAAGCAGAAA	CATATTCTTT	ACAATCACAA	1020
AAGACAACAT	TTTTTTCAAT	GCCATCCCAT	CTGTCTGTG	AAGAACTTCC	AATATATTTA	1080
TTTTTGCGTA	ATCTTTCCAT	CTCATATTGT	TTTTGAGGAG	CATATGGTTC	CCAATAATCT	1140
AATCCTTTTT	TTGTTCCAGA	ACGGCCTTTA	AGAACTTCTA	CATTTCTAGA	AGCTTTAATG	1200
TTATAATATG	AATAGATTAA	ACATTGTTTC	CCATCCACTT	CATCTATTTG	ATCAACATTT	1260
GTACTAAACC	AATATTCAGA	CACACTTTTA	TTGGCTGGAG	AACCATATAA	AGCTTGTAAG	1320
ATTGAAGTTT	TATTTACTCC	ATATCTATTA	CAGACACCTC	AGGATTATTT	AACTTATAAG	1380
TTTTAACAGC	TACCGAATCA	ATTTCAACAG	CAACTTGAAC	ATCTATGCCT	GATTTTTTAA	1440
GGCCACTTGT	AGTGCCACCT	GCACCGTTAA	ATAAATCAAT	AGCAACAATT	TTCCCCATAG	1500
TATTCTCCTA	AAGTTTCTCC	TTTTTATTAT	AACATTATCA	AATGTAAAAC	CCAACCCGAT	1560
AGGGTTAGGT	TTTTAACATC	ATTTACACAA	CTTCTTCATC	TCATCAATAC	GTGCGACGGT	1620
CGCGTCATAT	TTAGCTTGGT	AGTCAGCTTG	TTTGTGCGAT	TCTTTTTGGA	CGACTTCTGG	1680
TTTGGCGTTG	GCTACGAAGC	GTTTCGTTAGA	GAGTTTCTTA	CCAACCATGT	CCAGTTCTTT	1740
TTGCCATTTA	GCAAGTTCCT	TGTCGAGACG	GGCCAGTTCT	TCTTCAACAT	TGAGGAGATC	1800
GGCCAGTGGC	AGGTAGATTT	CTGCTCCTGT	GATGACACTT	GACATAGCCA	GTTCAGGTGC	1860
AGGGATGGTT	GATGCCATTT	CCAAGTGTTC	TGGATTTGTA	AAGCGTTTGA	TATAGTTGAC	1920
ATTGCTGTTA	AAGAAGGCTT	CCAAGTCGCT	ATCGCTTGTC	TTAACAAGGA	TGGTGATAGG	1980
CTTGCTTGGT	GCTACATTTA	CTTCCGCACG	CGCATTCCGA	ACAGCACGAA	TCAAGTCTTT	2040
GAGACTTTCC	ACACCAAGTGT	GAGCCGCAAG	GTCTTCAAAG	GCTAGATTAA	CAGTTGGGTA	2100
TGCAGCTGTC	ACGATAGAAC	CTTCTGAGAT	TTGTCCAAAG	ATTTCTCTCTG	TCACGAATGG	2160
CATGATTGGG	TGAAGGAGAC	GAAGGATCTT	GTCCAGCGTA	TAGAGGAGAA	CAGATCGAGT	2220
AATGACCTTA	TCGTCTTCAT	TGTCGCTGTA	TAGAACTTCC	TTGGTCAACT	CAACATACCA	2280
GTTGGCAAAT	TCTTCCCAGA	TGAAGTTGTA	AAGGATATGA	CCAGCCACAC	CAAACTCGAA	2340
CTTATCAAAG	TTTTTCAGTAA	CTTTTGCAAT	GGTTTCGTTG	AGATTGTGGA	GAATCCAGCG	2400
GTCCGTCACA	TTACCAGCCT	CACCTGTTGC	AACTTTTGTG	ACATTGTCTAT	GCGCCACATC	2460
CAGCGTCAAA	CCTTCATTGT	TCATGAGGAT	ATAGCGAGAA	ATGTTCCAAA	TTTTGTAAAT	2520

307

AAAGTTCCAT GAAGCATCCA TTTTCTCGTA AGAGAAACGA ACGTCTTGAC CTGGTGCGGA	2580
ACCGTTTGAA AGGAACCAAC GAAGGCGATC AGCACCGTAT TTCTCGATGA CATCCATTGG	2640
GTCAATCCCG TTACCGAGAG ATTTAGACAT CTTGCGTCCT TGCTCGTCAC GGATGAGACC	2700
GTGGATAAGC ACGTTTGGGA ATGGCTGACG ACCAGTAAAT TCCAAGGACT GGAAGATCAT	2760
ACGAGACACC CAGAAGAAGA TGATGTCGTA ACCTGTTACC AAGGTGGAAG TTGGGAAATA	2820
ACGTTTAAAG TCTTCTGAGT CGACTTCAGG CCAGCCCATG GTTGAAAATG GCCAGAGGGC	2880
AGAACTGAAC CAAGTATCCA AGACGTCTTC GTCCTGAGTC CATCCGTCAC CTTCTGGAGC	2940
TTCTTCGCCC ACATACATTT CACCATCAGC ATTGTACCAG GCAGGGATTT GGTGACCCCA	3000
CCAAAGCTGA CGAGAGATAA CCCAGTCGTG GACATTTTCC ATCCATTGAA GGAAGGTATC	3060
GTGAAACGA GGTGGGTAGA ATTCGACCTT GTCCTCTGTG TCTTGGTTAG CAATGGCGTT	3120
CTTAGCCAAT TGGTCCATCT TGACGAACCA TTGAGTAGAC AAGCGTGGCT CAACTACGAC	3180
ACCTGTACGT TCTGAGTGAC CAACACTGTG GACACGTTTT TCGATTTTGA CAAGGGCACC	3240
GATTTCTTCC AACTTAGCAA CGACTGCTTT ACGAGCTTCA AAACGATCCA TGCCTGAAAA	3300
TTCAAAGGCA AGCTCATTCA TAGTTCCGTC GTCGTTTCATG ACGTTGACTT GTGGCAAGTT	3360
ATGACGTTGG CCAACCAAGA AGTCATTTGG ATCGTGGGCA GGTGTGATTT TCACGACACC	3420
AGTACCAAGC TCAGGATCTG CGTGCTCATC TCCAACGATT GGGATGAGTT TATTAGCGAT	3480
TGGAAGGATG ACGTTTTTAC CAATCAAGTC CTTGTAGCGC GGGTCTTCTG GATTAAACGC	3540
AACCGCAACG TCCCCAAACA TAGTCTCAGG ACGAGTTGTA GCAACTTCAA GGGCGCGTGA	3600
ACCATCTTCC AGCATGTAAT TCATGTGGTA GAAGGCACCT TCTACATCCT TGTGAATCAC	3660
CTCAATATCA GAAAGGCGTG TGGAGCTGC TGGGTCCCAG TTGATATAA ACTCAGTACG	3720
ATAGATCCAG CCTTTCTTGT AAAGGTTTAC AAAGACCTTA CGAACAGCTT TTGACAAACC	3780
TTCATCAAGA GTGAAACGCT CACGAGAATA GTCTACAGAA AGCCCCATCT TGCCCCATTG	3840
TTCTTGATG GTAGTGGCAT ATTCGTCTTT CCATTCCCAG ACCTTCGTCA AGAAAGACTC	3900
ACGACCTAGG TCATAACGCG TAATACCCTC ACCACGTAAG CGCTCCTCAA CCTTAGCCTG	3960
AGTCGCAATA CCAGCGTGGT CCATACCTGG AAGCCAAAGG GTATCAAAGC CTTGCATGCG	4020
TTTTTGACGG ATGATGATAT CCTGCAAAGT CGTATCCCAA GCGTGACCAA GGTGAAGTTT	4080
CCCAGTTACG TTTGGTGGTG GAATCACGAT TGAATAAGGC TTAGCCTTTT GATCGCCTGA	4140
AGGCTTGAAA ACATCCGCAT CAAGCCATTT TTGGTAACGA CCAGCCTCAA CCTCGGCTGG	4200
ATTGTATTTA GGTGAAAGTT CTTTAGACAT GTGTGTGTCC TTTCTCTATT TTGTTTATTT	4260

308

TATTTTGAAT TTGCTTAGCA GCTTCTTCTG CAGACAAATT CGTATTATTT ATTTTAAAGT	4320
AGTGGTGCAA CTCATTCCGT TGATGTTGGG AATTTAATTG AAGTGTTTCA GCGGTCTCTA	4380
AAATTTCCTT TTCAGATACC TCAATATGTC GTTTTAAGGG TTTGTGCTTT AATCGATTCT	4440
CCGTTTCGATT TCGACGTATG CACTCTTCAA GACTTGTTTC CAATTCAACA AACAGAATCT	4500
CTTGATGAAA GTTATCCAAT AAATCCTGAA TTTGCTTTAA ATACATCAGC TGGTACTGAT	4560
TTGAAAAATC AATTACGTCT GTTAAAAATTA CTGATCGCTG ATTTCTTGCA CTTGCTCCAA	4620
GGAAAGAAAA GGTAATTCCA CGAACAAATT CCCACATCTC CTCGGTATAA TCCTGATAGA	4680
TCTCTAGTGC AAAATCAATG GCTTGATGGT TATAAAATAG GGTAGCATCC GTCAGTCGAG	4740
ATAATTCTTG ACCAATGGTC ATTTTTCCTG ATGCTGGAGC ACCAATGATG AAAAGATGCA	4800
TCAAATCACC TCCCCTCAC TCCTCAGCAA GCCATATCTC AAATCATCAC AGCAGTTGCC	4860
TTGAGCATCT TTGCGGTCTC TTATGCGAGC TTCGAGGGTA AAGCCAAGCT TTTCCGAGAC	4920
TCGTTGACTT TGAAGGTAT ATCCAAAGCA AGTTAGTTCA ATCTTGTAAG GACCAAGTTC	4980
TTTAAAAGCT AGATCAATCA AGGAACACGC TGCTTCTGGA ACATAACCTC GACCCCAATA	5040
GTCTGGGTGC AAGGTATAGC CAAGCTCTAG CACATCATCC GCATGAAGAT GGTGAAGTC	5100
AACAGAACCA ATGACTTTAT CGGTTCTTTT GACGACAATC CCATAGCCAG CTGGGAGATT	5160
TTCTTTTGA GTACGCTCCG GAAGAATGTG CTCCAGATAA TAAATCTCAT CTTCCAAGAT	5220
CTTGACTGGA GGAAAACCTG CTGGATAGGC GACCTCTGGC AAAGTAGCGT AGGTATGGAT	5280
ATCCTCAGCA TCCACCACTG TGCGGACTCG TAAAACGAGA CGTTCTGTTT CGATTTTATC	5340
TGGCAGCTCA GTTCTTGCCA TCCTTCTTCC TCGCTTTTTT GATGAAACTG CCCTTCATAT	5400
CTACACGCTT GTCCAGATAG CGATAAACGC GCTGATATCC ATCTCCCATG AAATAGGTTG	5460
GGGCAAACAG TTGATTTTTA AAATGTCCCT TTTCATCCAG GAGTCTGGG GCAACAAGTC	5520
GCTCAAGAAT CTTGGCAAAG ATGTGGCAAA TACCGTCTTC CTCACAATC CTATCTACCC	5580
GACAACTAA AACAAAGTGA CAGGCGTCTA AAATAGGAGT CTGAGTTCGT TCAGAAATTT	5640
CATAATGCAC TCCCAAACGT TCCAATTTCT CCTGATGACT GATAAAACCA GCCTGCTCCA	5700
TCGCAAGCAT AGAAGTTTCA TCAGAAATAT TCACAGTAAA TTTTGTATAC TGTTTGATCT	5760
GCTCTGCGGC ATTCTCTCTC GCAACGACTC CAATCACAAC CCAATCTCCT AGACTATAAG	5820
AGGAACTACA GGTCGTGATG TTATAGCCAA AATTCTAATC TTGATATCCT AAAATAAAAA	5880
CAGGAAAACC ATAATATAGT TTAATTGTGT TAAAAGATTG CTTTATAACA ACCCCCTTTG	5940
ACTAAGACGT AAAAGAAAAG CCCTGCCATC TACATGACAG GGACGAATGT GTTTATCCGC	6000
GGGG	6004

309

## (2) INFORMATION FOR SEQ ID NO: 28:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5857 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

TGTAGAATTC ACGACAATGC TTCGTTGATT TCTGGGTGA TTTCGTCGCG TTCTGGCAAG	60
CGAGTCAATG AACCAAAAAT AGTACACAAT GTGGTATAAT CCTTTTATGG CATATTCAAT	120
AGATTTTCGT AAAAAAGTTC TCTCTTATTG TGAGCGAACA GGTAGTATAA CAGAAGCATC	180
ACACGTTTTTC CAAATCTCAC GTAATACCAT TTATGGCTGG TTAAAGCTAA AAGAGAAAAC	240
AGGAGAGCTA AACCAACAAG TAAAAGGAAC AAAACCAAGA AAAGTTGATA GAGATAGACT	300
TAAAAACTAT CTTACTGACA ATCCAGATGC TTATTTGACT GAAATAGCTT CTGACTTTGG	360
CTGTCATCCA ACTACCATCC ACTATGCGCT CAAAGCTATG GGCTACACTC GAAAAAGAA	420
CCACACCTAC TATGAACAAG ACCCAGAAAA AGTAGCCTTA TTTCTTAAGA ATTTTAATAG	480
TTTAAAGCAC CTAACACCTG TTTAGATTGA CGAAACAGGA TTCGATACTT ATTTTATCG	540
AGAATATGGT CGCTCATTAA AAGGTCAGTT AATAAGAGGC AAAGTATCTG GAAGAAGATA	600
TCAGAGGATT TCTTTGGTTG CAGGTCTAAC AAATGGTGAG TTAATCGCTC CAATGACTTA	660
CGAAGAGACG ATGACGAGCG ACTTTTTTGA AGCTTGGTTT CAGAAGTTTC TCTTACCAAC	720
ATTAACCACA CCATCGGTTA TTATTATGGA TAATGCAAGA TTCCATAGAA TGGGGAAGCT	780
AGAACTCTTG TGTGAAGAGT TTGGGTATAA ACTTTTACCT CTTCTCCCT ACTCACCTGA	840
GTACAATCCT ATTGAGAAAA CATGGGCTCA TATCAAAAAG CACCTCAAAA AGGTATTACC	900
AAGTTGCAAT ACCTTTTATG AGGCTTTTTT GTCTTGTCT TGTTCATT GACTATATAA	960
ATTGTCTAAG CGAAACAACC GATAAGAATT GGCACAAAAG CGACCGTATT TTTGTTACCA	1020
ATACAGGAAA AACAGTTCAT AGTCTATCT TGAGCAAGTC TCTCCAGCGA GCAAACGAAC	1080
GCCTTAAAAA ACCAATTCCT AAACATCTGT CCCCTCACAT CTTCAGACAC ACCACTATTA	1140
GCATCTTATC AGAAAATAAA ATTCCTTTAA AAACAATCAC GGACAGGGTT GGTATCCCCG	1200
ACTCTGAAGT CACTACTTCC ATCTACACCC ACGTCACAAA GAACATGAAA GATGAAGCAA	1260
TCAATGTACT GGATAAAGTT ATGAAAAAGA TTTTAAAAA AGTTTGTCC CTTTTTGCC	1320
CTCTAAATAC AAAAATAGCC CTTCCGATAA AATCCGAGGG GCTAGAAACG TTGTTAAATC	1380

310

AACGGCCGAA	CTTTTGAATT	TCATGGTTCG	GGATAAAATA	GTTCACGTAA	CTATTTTATT	1440
TTTTAAGGTT	ATCATAATAT	CAAATAGTTC	AATTAAATAC	GCTAAATTAC	TAATATACTT	1500
TTTACCTTTT	TCATTCTAAA	ATGTAAAGTA	CAAACAATTA	CAATATACTA	GAGGGGGAGT	1560
AAAAAAGGTA	TTAAATCGAT	GAGTTCAGCA	GGCAAGAAAA	TAGCACCTTT	ACGGGTGCTA	1620
TTTTTTAATT	AACGCCACGT	TAACTTTTGA	TTGATGAATT	TTATTGTTTG	GCACCTCTTT	1680
CATTTACCGG	TAAACATCGA	TGAAATTCTT	TCCAACATTA	TTTTTGGAGT	TAAGTGCATT	1740
TATTTTTGTA	TTAATAACTT	TTTAGTATC	GAAAGAATGG	TTTAAGAAAT	CCATAACTAA	1800
CTCTCCTTTC	TCATCCTGTA	ATCAAGATTT	TTATCAATGT	CAAAATAGTA	TTTTCTATCA	1860
ATCCAAATTG	GTCTTCTCC	TTTAGAAATA	GCAAGTACAT	CTACCGGACC	TCCTACTGTT	1920
TCAAGAGTGT	TGACAATTTT	TCTCTTAAAT	GAAGTTAATT	CAATAAATGT	TTTAGCTGTA	1980
CTCGCCATTT	CATTAAGTGG	TTGCATTCCA	ATAAGGTCTA	TTATAGGATT	TATATAATAT	2040
TTTTGCTGTA	TAGATGATAT	ATTTTCAAAT	ATATTCTCAA	TTTCATCACC	CAATCCATTT	2100
TTCTCCATAA	CTGATGATAC	TTGCTCTGCG	ATATATACAT	TTAAGTTAGG	ATCTATACCA	2160
TTCATAATCG	TCTCAACCAT	CTCTGACTGT	GCAAAAGGGA	TTATATGACA	AGTTTTATGA	2220
TGATTTATCA	CACTTTCATT	AATAACTTTC	CAAATTAATC	GTTTAGAAAA	AATTCCATAT	2280
AATTCAATTT	GTCTTATAGA	TGGAAATATC	TCGTCTGTAC	CATAACCTGC	TATAACTAAT	2340
CCAGTTATGT	TTGTTGAGTC	ATATCCAATG	AAAATCGCTT	TATATAAAGA	TTTAGCAATA	2400
ACTTCAACCT	CATCATCAGT	ATGAGGAAAG	GATTTAAAAA	CATCGTCTAC	AATGCTTTTT	2460
ATTAACCTTA	ACTCAGCTTC	AAAAAATTC	AAATTACTTT	CAGCTTCTAC	TTTTGAAATT	2520
TCTAAACTAA	AATTAGTTAT	AGCATTTAAT	AAAATTTTAT	TAAAATCATC	TAGAGTGATG	2580
GTTTCACCAT	TAGAACTCT	TAAATCAGCT	GTTTCTTGCG	CTTCATAGGC	AATGCTGTCC	2640
AAAATACTTC	TTGTACTTCT	GACAAATATA	TTCTTAATA	AATCCTCAAC	TTGTAGATGT	2700
TTAAAGGAAA	TTAAAAATTC	TATTAGCTTT	TCAACGTATT	GGGCAGTATT	ATCTAATAAA	2760
TCTGTGCCAA	TAGCCTGCTT	AAACTCATTT	AAAATTACCT	CCCACGGAAT	TTCCATAAAC	2820
GAAGCGTTCC	CATATATCAT	GATCCCCACG	GAATGTTCTT	TTGATAAAGT	GAATAATTTT	2880
CGGGCGCTAT	TAAAAACTTT	TGAATTTTTC	CCGTCTGATA	AGGTTACAGC	GCTATCAGAA	2940
GCCAATACAA	CACCATTTTT	ATTTAATATT	CCAATTTCTG	CTGTCAAAAT	ATCACCTAAA	3000
CTTCTAAAC	CTGCTCATGC	TCTAATGGTA	CAACAGCTAA	GGTCTTACCA	AGACTTGCCA	3060
ACACTTTTAA	TACTGTATCA	AGTTGTGGGC	TTGTCTTTCC	TGTTTCCATT	CTAGCGATAA	3120
CTGGCTGACT	AACACCGCTC	ATCTCCTCTA	GTTTCTTCTG	ACTAATACCC	TTTTCATTTT	3180

TAGCCTCGAT AAGCTCACTC ATGATAGCCA CGCGCATATC ACTTTCCAAA ATTTCTCTTT	3240
TGCTGAATAA TTCAGCTCTT ACATCTTTCC AGTTACTACC AATAGCATT ATTTTTCATTG	3300
TCTAAACCTC TTTCTTTTAA ATCTGCAAGT TCACGTTTAG CTTGCTCAAT CTCTCTTTTG	3360
GGTGTTTTCT GTGTCCTTTT CATAAAATGA TGCAGTAAAA CAAAACCTACC ATCCATCCAA	3420
GCAACAAATA AAATTCTATC TCTAAGTGGT CTCAGCTCCC AAATTTCAAGC ATCTAAATGC	3480
TTAATATATG GTTCGCCTGC GCGTGTCCA TGTGGGCTTA ACAACTCAAT ATAATCATT A	3540
ATTTTATTAA GCTTAATTCT GCTATCTTTC CCTTTTTTAC TGGTAAGCTC TCGCATATAA	3600
TCAAAAACAG GCTCATTGCC GTTTTATATCC TTGTAAAAAT AGATATTATG CACTATTAAC	3660
ACCTCTTCCT AATAACAATT ATAACCTAAA AGTTATTGTT TGTAAATACT TTTAAGTTAT	3720
TAAAATAAAA AGCACCTAGT TTCCTAGATG CTAGCACAAT GACACGGATT CGCACCGTGG	3780
CTACCTCTAT CAAGGTGTAC TCCTTCTATA CTATCCCTTG TGCTTTAGAA TATTATACCA	3840
CACAATCAAC TAGATACCTA CCATCTCATG ATATACCCCC ATTTTGGGCA AGGGTACAAC	3900
GCTAAAATAC AAATCAGAAT AGATATTAAA CCACTTATTT AACTTATCAT AAGCTGGTGA	3960
TTGACTGATA AATAATATCC GCTGACAAGC TCCGATAACA TTCATGTGAT TGTACACATA	4020
AACCTCTTTT ACAGCCTCTA AAATGTCAGC CTCACTTGGT TGTACCCTAA TATCTGTTAT	4080
CTGCTTGATA GTTGCCTATT TTTGATAAGC TAGCATATCT TGATTTTTAG CAGCATCAAA	4140
CATTTTACGC TCAAGGACAC TATACTTAGG TTGTTCTTTA TCTCGCATGA AATACCCTT	4200
GAGCCATAAA ATCTTTTCTC GGTGTATTAC AGAAATACGC TCAATTTTCT TCTTTGTCT	4250
TGCTACCTCC TAAATCATCA ATTTAACAAT TCTAACCCT CACTTTTAGA AATAGTTGCA	4320
TAGATCTTGT TCGATGTATG ATACAAAGGT TCTAAATCTT TTTCCACCCT AATATAGTTC	4380
ATCTTATCCT CATGAGTAGG AAAGTATAGT ATTTCCGTTT CATCCTCGTT TAGGATACGA	4440
TTGCACCAAT CATCAATAAT AACTGGCACT TCCCACTCAC GCCATTTTTT AAGGTTTTCT	4500
AAAAGTTCAT TATCACTAAA TAGCTCGCCA TCTATTTGGA AAAATTCCCC TAAGTCATTG	4560
TTTCCTTCAA CAATAATAAA CTCTGGCATA TTTCTATTAC TTAATAACTC CTTGAGTTCT	4620
TGTAACCTCT TGATTTCTTT TAGATACTTC CTCAATTTCC AACCTCAATT CTTCAATCTG	4680
CCTTACTACT CCAAAAATTT CATGGGTCTT ATAAGATTGT TCAAGTATAG CCTTTGCTGC	4740
TTGAGTTCTT ATAAACGGGT TGACCTTACT GTCCATCATA ATATCATTGA GTACAGAAAC	4800
AGCGTTAGAT GATGCTAAAT AAAGCATTTG AGTTGTTTTA TCCATCATCT CATCTTGCTT	4860
TATCCTCAAT GTCTTTTAA CCGCTGCAAC TTTTAGATAC TTATGACCTG TTGCGCGTGA	4920

312

TACCCCTGCT TTTGACATG CTTGTCTAT CGTTGGCTCG GTAAGCATGG CATCTATGAA	4980
TTTAATTTGC TTGGACGTAA GGTATCATT TTCATTTCTT GCCATCTATT ACCTCCTCAT	5040
TATCAAAATA AAGGGTTGCC CCTTTATTTT CCTATGCTAG ATAATTCTGC AATTCTGCAT	5100
CCATTGCCTC TGAATTGCCC TCAACAATCA TTTCATGCTG TACTAAATCA ATCTTATCTC	5160
CGTTAATAAG TAAACCACCG TGGAAATAAT CAATTTTTCT ATCAAGGAAA TGTACTAGCT	5220
TTTCAAGGCG TTGCTGTTGG CTGAATTGCT CCATGTCAAT TTCGATATAA GCAAGGGTAG	5280
TATCATTATC CATAATATCT TCTAATTTTC TAAGAGCTAG AGGTTTATTT TTATATTTTT	5340
CTAGGTATTC TCTCATTTCT GCCACTGTTA ATTTGATACT AGATAATAAA CTTAGTTCAG	5400
CTGCATCATC TGCTGTAATA GGCTCTTCTT TTGATTCAAT GTTTGCTAGT TCAGCATTTT	5460
TCTCTTTTTT TAGTTGCTGA TACAATAGCT GAGCAGTATT TTGGGAATAG TTTTCGCCCT	5520
CTTTTTTATA TTTTAAAAGT TCTTGCTCTG CATACACTTT CCGGATAATC ACTTCCTTAT	5580
AAACTAATCG CCCATCTTGA GCTTTTAGCT TAATACTCCC ATGCTCTGGA ATTTCAATAT	5640
ACTTAATTAT ACCATTTTTT GAGTATAAAA CAAAGCCTTT CTCCATCATT TTTAATAATT	5700
TATCATCCTT GTTTTCAGTC ATGCTTTTCT CTTTTATTTC ATTTTATTAT AATCTGAATA	5760
CCCCTAGTCT ATTTATTTCA CTAGGTTTTT AGGGTTCGTA TGCTAAAATA CTACCCTTTT	5820
TGTGTACCTT ATGGCTGACT TTTCAAATTG GTTAGTT	5857

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10254 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AAAATGATAG CAGGAGAGTT TTCCCGTCCA TCAGACCCAG AACTGAGAGC CTTAGCTCAG	60
GCTTCTCGCC AAAACAGGC CGCCTTTAAC AAGGAAGAGA ACCCCTTGAA GGGAGCCGAA	120
ATCATCAAGA CTTGGTTTGC CTCAACCGGG AAAAATCTTT ACATCAACAC TCGCTTGATG	180
GTGGACTACG GTGTCAACAT CCATCTAGGG GAAAATTTTT ATTCTAATTG GAACTTGACC	240
ATGCTGGATA TCTGTCCCAT TCGTATCGGG GACAAATGCTA TGATTGGTCC TAATTGTCAG	300
TTTTTGACAC CCCTCCATCC ACTAGATCCA CAGGAACGCA ATTCAGGTAT CGAGTACGGA	360
AAGCCTATCA CAATCGGAGA TAATTTCTGG ACTGGTGGTG GCGTCATTGT CCTTCCTGGA	420
GTGACACTGG GAAATAATGT CGTTGCAGGA GCAGGGGCAG TAATTACCAA ATCTTTTGGC	480

313

GACAACGTTG TCCTAGCTGG CAATCCTGCG CGCGTGATTA AGGAAATACC TGTAAATAG	540
AAGTAAAAAG GAACAGCTGG GGTGTTTCT TTTTGTAGG TTTCATCATT TTTTACCCAG	600
TTACATTTA CCTACTCTAT CTCTTAGCAA GTCTGTTTCA TTAAGCAAGT TCAAAGCATC	660
TCGTAAGTGG GATGTTTTTC TCCTCAGTTC ATCAGCTTCC TCCTTGACAC TCGGTCAGAT	720
TTTGATACAA TAGTACAAAA TTAGAGGAGG CAGGCTATGA TTCAGAAACA TGCGATTCCT	780
ATTTTAGAGT TTGATGACAA TCCTCAGGCG GTTATCATGC CCAATCACGA GGGGCTGGAC	840
TTGCAGTTGC CAAAGAAGTG TGTTTATGCA TTTTtagGTG AGGAGATTGA CCGCTATGCG	900
AGGGAAGTAG GGGCGAAGTG TGTGCGGAA TTTGTTTCTG CCACCAAGAC CTATCCAGTT	960
TATGTCGTGA ACTACAAGGA CGAGGAGGTC TGTCTGGCTC AGGCTCCTGT TGGCTCCGCT	1020
CCAGCAGCCC AGTTTATGGA TTGGTTGATT GGCTATGGTG TGGAGCAGAT TATCTCTACT	1080
GGGACCTGTG GTGCTTAGC TGATATAGAG GAAAATGCCT TTCTAGTCCC TGTTCGCGCT	1140
CTGCCAGATG AAGGAGCCAG TTACCACTAT GTGGCACCTT GTCGTTATAT GGAAATGCAG	1200
CCAGAGGCTA TTGCTGCTAT TGAGGAAGTT TTGGAAGACA GAGGGATTCC TTATGAAGAA	1260
GTCATGACCT GGACGACAGA CGGTTTTTAC CGAGAAACGG CTGAAAAGGT GGCTTATCGT	1320
AAGGAAGAAG GCTGTGCTGT TGTGGAGATG GAGTGTCTG CTCTTGCGGC AGTAGCTCAA	1380
TTGCGTGGGG TTCTCTGGGG TGAATTGTTG TTCACAGCAG ATTCTCTAGC GGAATTGGAC	1440
CAGTACGACA GTCGTGACTG GGGCTCGGAA GCTTTTAATA AGGCGCTAGA ACTGAGTTTA	1500
GCAAGTGTTC ACCACCTTTA GTTGTAAGTG CAAAGGATTT GTTTTATCAT AAAATGTCTA	1560
GCTCATACTT TTCAAAAATA TGTTTAAACG AGGTCACCTT CCTCTGTGCC TAGGCATGTT	1620
GAGGTGTTGA AAAATCTTTA AAATCAGAAA AACGTATCAT ATCAGGTGAT GAAAACCTTG	1680
AACTATGCG TTTTATGTG ATAAGATTTA GAGTGAGATG AAATGATACT CTTGAAAAAT	1740
CTCTTCAAC CAGGTCAGCT TCACCTTGCC GTAGGTATAT GTTACTGACT TCGTCAGTCT	1800
TATCCGGCAA CCTCAAAACG GTGTTTTGAG CTGACTTCGT CAGTTCTATT TGCAACCTCA	1860
AAACAGTGT TTGAGCAACC TGTGACTAGC TTTCTAATCG ATGCCCTGGT TTTTATTGCC	1920
TATAATCAAA AAGAGAAATT TTCTCCTGAA AAGCATATAG AGTAGCTGGC GTTAAAAGCT	1980
CCTGTCTTG TTTTGTGACC TATAGTCACA TCTATCAAGT ATTGTTCTTG CCTAAGCTAT	2040
CAATAAAAAG GTGGCATTTT TTAGGCTTGG TGTTAGTAGA TTTTGCCTTA TCCTATCTAA	2100
GTCATTTTGA ACTTTTTATG GTACAATGGA AACATGTTAT TCAAATTATC TAAGGAAAAA	2160
ATAGAGCTAG GCTTATCTCG TTTATCGCCA GCCCGTCGTA TTTTTTGAG TTTTGCCTTG	2220



314

GTCATTTTAC TAGGCTCTCT TCTTTTGAGC TTGCCCTTTG TCCAAGTTGA AAGCTCACGA	2280
GCGACTTATT TTGATCATCT TTTCAC TGCTCTGCAG TCTGTGTGAC GGGTCTCTCA	2340
ACCCCTCCAG TAGCTCACAC CTATAATATC TGGGGTCAAA TAATCTGTTT GCTCTTGATT	2400
CAGATCGGTG GTCTAGGGCT CATGACCTTT ATTGGGGTTT TCTATATCCA GAGCAAGCAA	2460
AAGCTTAGTC TTCGTAGCCG TGCAACTATT CAGGATAGTT TTAGTTATGG AGAAACTCGA	2520
TCTTTGAGAA AGTTTGCTTA TTCTATTTTT CTCACGACCT TTTTGGTGGA GAGCTTGGGA	2580
GCTATTTTGC TTAGTTTTCG CCTTATTCCT CAACTTGGCT GGGGACGTGG TCTTTTTAGT	2640
TCCATTTTTC TAGCGATCTC AGCCTTCTGT AATGCCGTTT TTGATAATTT AGGGAGCACC	2700
AGTTTATTTG CTTTTCAGAC CGATTTACTG GTCAATCTGG TGATTGCAGG CTTGATTATT	2760
ACAGGCGGCC TTGGTTTTAT GGTCTGGTTT GATTGGCTG GTCATGTAGG AAGAAAGAAA	2820
AAAGGACGTC TGCACCTTCA TACGAAGCTT GACTATTAT TGAATATAGG TTTGTTGTTA	2880
TTTGGACAG CAACTACTCT CTTTCTTGAG TGGAAACATG CTGGAACGAT TGGCAATCTC	2940
CCTGTTGCCG ATAAGGTTTT AGTTAGCTTT TTTCAAACAG TGACGATGCG AACAGCTGGC	3000
TTTTCTACGA TAGATTATAC TCAGGCTCAT CCTGTGACTC TTTTGATTTA TATCTTACAG	3060
ATGTTTCTAG GTGGGCGACC TGGAGGAACA GCTGGGGGAC TCAAGATTAC GACATTTTTT	3120
GTCCTCTTGG TCTTTGCACG AAGTGAGCTT CTAGGCTTGC CTCATGCCAA TGTGCGAGA	3180
CGAACGATCG CGCCGCGAAC GGTTCAAAAA TCCTTTAGTG TCTTTATTAT CTTTTTGATG	3240
AGCTTCTTGA TAGGATTGAT TCTGCTAGGG ATAACAGCCA AAGGCAATCC TCCCTTTATC	3300
CACCTCGTAT TTGAAACCAT TTCAGCTCTT AGTACAGTTG GTGTAACGGC AAATCTGACT	3360
CCTGACCTTG GGAAATTGGC TCTCAGTGTT ATCATGCCAC TTATGTTTAT GGGACGAATT	3420
GGTCCCTTGA CCTTGTTTGT TAGCTTGGCA GATTACCATC CAGAAAAGAA AGATATGATT	3480
CACTATATGA AAGCAGATAT TAGTATTGGT TAAGAAAGGA AAGAGCATGT CAGATCGTAC	3540
GATTGGAATT TTGGGCTTGG GAATTTTGG GAGCAGTGT CTAGCTGCCC TAGCCAAGCA	3600
GGATATGAAT ATTATCGCTA TTGATGACCA CGCAGAGCGC ATCAATCAGT TTGAGCCAGT	3660
TTTGGCGCGT GGAGTGATTG GTGACATCAC AGATGAAGAA TTATTGAGAT CAGCAGGGAT	3720
TGATACCTGC GATACCGTTG TAGTCGCGAC AGGTGAAAAT CTGGAGTCGA GTGTGCTTGC	3780
GGTTATGCAC TGTAAGAGTT TGGGGGTACC GACTGTTATT GCTAAGGTCA AAAGTCAGAC	3840
CGCTAAGAAA GTGCTAGAAA AGATTGGAGC TGACTCGGTT ATCTCGCCAG AGTATGAAAT	3900
GGGGCAGTCT CTAGCACAGA CCATTCTTTT CCATAATAGT GTTGATGTCT TTCAGTTGGA	3960
TAAAAATGTG TCTATCGTGG AGATGAAAAT TCCTCAGTCT TGGGCAGGTC AAAGTCTGAG	4020

TAAATTAGAC CTCCGTGGCA AATACAATCT GAATATTTTG GGTTCGAG AGCAGGAAAA	4080
TTCCCCATTG GATGTTGAAT TTGGACCAGA TGACCTCTTG AAAGCAGATA CCTATATTTT	4140
GGCAGTCATC AACAACCACT ATTTGGATAC CCTAGTAGCA TTGAATTCGT AAAGAGGGAT	4200
GACCCCTCTT TTTTGATGCC TAAGATGGCA AATAGAGACA GAAGCCCTT GTCTTCTAGT	4260
AAAAGTTCTT CAAAGGCTGG ACTTTATGGT AAAATAGAAA GAAGTGACAA GAGAGAGTAA	4320
TACTCAATGA AAATCAAAGA TCAAACCTAGG AAACCTAGCTA CGGGCTGCTC AAAACACTGT	4380
TTTGAGGTTG CAGATAGAAC TGACGAAGTC AGTAACATCT ATACGGCAAG GCGACGTTGA	4440
CGCGGTTTGA AGAGATTTTC GAAGAGTATA AGAAAAATC AGTCCCTAA AGGAGTAGAT	4500
TATGAAGTTA TTGTCTATCG CAATTTCTAG CTATAATGCA GCAGCCTATC TTCATTACTG	4560
TGTGGAGTCG CTAGTGATTG GTGGTGAGCA AGTTGGGATT TTGATTATCA ATGACGGGTC	4620
TCAGGATCAG ACTCAGGAAA TCGCTGAGTG TTTAGCTAGC AAGTATCCTA ATATCGTTAG	4680
AGCCATCTAT CAGGAAAAATA AATGCCATGG CGGTGCGGTC AATCGTGGCT TGGTAGAGGC	4740
TTCTGGGCGC TATTTTAAAG TAGTTGACAG TGATGACTGG GTGGATCCTC GTGCCACTT	4800
GAAAATCTT GAAACCTTGC AGGAACTTGA GAGCAAAGGT CAAGAGGTGG ATGTCTTTGT	4860
GACCAATTTT GTCTATGAAA AGGAAGGGCA GTCTCGTAAG AAGAGTATGA GTTACGATTC	4920
AGTCTTGCCCT GTTCGGCAGA TTTTGGGCTG GGACCAGGTC GGAAATTTCT CCAAAGGCCA	4980
GTATACCATG ATGCACTCGC TGATTTATCG GACAGATTTG TTGCGTGCTA GCCAGTTCTA	5040
ACTGCCTGAA CATACTTTTT ATGTCGATAA TCTCTTTGTC TTTACGCCCC TTCAGCAGGT	5100
CAAGACCATG TACTATCTGC CTGTCGATTT CTATCGTTAT TTGATTGGGC GTGAGGACCA	5160
GTCTGTCAAT GAGCAAGTGA TGATTAAGTG CATTGACCAG CAACTCAAGG TCAATCGACT	5220
CTTGATAGAC CAACTTGATT TGTCCCAAGT GAGTCATCCC AAAATGCGAG AATATCTGCT	5280
GAATCATATT GAACTCACGA CGGTGATTTT CAGTACCCTG CTCAACCGAT CTGGAACAGC	5340
GGAGCATCTG GCAAAAAAAC GCCAATTGTG GACCTATATT CAGCAGAAAA ATCCAGAAGT	5400
CTTTCAGGCT ATTCGTAAGA CCATGTTGAG CCGTTTGACC AAACATTCTG TCTTGCCAGA	5460
TCCGAACTG TCCAATGTGG TCTATCAAAT CACCAAATCT GTTTATGGAT TTAATTAATA	5520
TAAGTGTTTT ATAAGAGGGA TTTAAGAAAA ATTTAACTT TTTCTTAGTC CTTTTTAATT	5580
TCAGGAGATT ATACTAGAGT CATCAAATAA AGAAAGACTC TAAGGAGAAT CCTATGAAAT	5640
TCAATCCAAA TCAAAGATAT ACTCGTTGGT CTATTCGCGG TCTCAGTGTC GGTGTTGCCT	5700
CAGTTGTGT GGCTAGTGGC TTCTTTGTCC TAGTTGGTCA GCCAAGTTCT GTACGTGCCG	5760

316

ATGGGCTCAA	TCCAACCCCA	GGTCAAGTCT	TACCTGAAGA	GACATCGGGA	ACGAAAGAGG	5820
GTGACTTATC	AGAAAAACCA	GGAGACACCG	TTCTCACTCA	AGCGAAACCT	GAGGGCGTTA	5880
CTGGAAATAC	GAATTCACCT	CCGACACCTA	CAGAAAGAAC	TGAAGTGAGC	GAGGAAACAA	5940
GCCCTTCTAG	TCTGGATACA	CTTTTGTAAA	AAGATGAAGA	AGCTCAAAAA	AATCCAGAGC	6000
TAACAGATGT	CTTAAAAGAA	ACTGTAGATA	CAGCTGATGT	GGATGGGACA	CAAGCAAGTC	6060
CAGCAGAAAC	TACTCCTGAA	CAAGTAAAAG	GTGGAGTGAA	AGAAAATACA	AAAGACAGCA	6120
TCGATGTTCC	TGCTGCTTAT	CTTGAAAAAG	CTGAAGGGAA	AGGTCCTTTC	ACTGCCGGTG	6180
TAAACCAAGT	AATTCCTTAT	GAATATTTCG	CTGGTGATGG	TATGTTAACT	CGTCTATTAC	6240
TAAAAGCTTC	GGATAATGCT	CCTGGTCTG	ACAATGGTAC	TGCTAAAAAT	CCTGCTTTAC	6300
CTCCTCTTGA	AGGATTAACA	AAAGGGAAAT	ACTTCTATGA	AGTAGACTTA	AATGGCAATA	6360
CTGTTGGTAA	ACAAGGTCAA	GCTTTAATTG	ATCAACTTCG	CGCTAATGGT	ACTCAAACCT	6420
ATAAAGCTAC	TGTTAAAGTT	TACGGAAATA	AAGACGGTAA	AGCTGACTTG	ACTAATCTAG	6480
TTGCTACTAA	AAATGTAGAC	ATCAACATCA	ATGGATTAGT	TGCTAAAGAA	ACAGTTCAAA	6540
AAGCCGTTCG	AGACAACGTT	AAAGACAGTA	TCGATGTTCC	AGCAGCCTAC	CTAGAAAAAG	6600
CCAAGGGTGA	AGGTCCATTC	ACAGCAGGTG	TCAACCATGT	GATTCCATAC	GAATCTTTCG	6660
CAGGTGATGG	CATGTTGACT	CGTCTCTTGC	TCAAGGCATC	TGACAAGGCA	CCATGGTCAG	6720
ATAACGGCGA	CGCTAAAAAC	CCAGCCCTAT	CTCCACTAGG	CGAAAAACGTG	AAGACCAAAG	6780
GTCAATACTT	CTATCAAGTA	GCCTTGGACG	GAAATGTAGC	TGGCAAAGAA	AAACAAGCGC	6840
TCATTGACCA	GTTCCGAGCA	AAYGGTACTC	AACTTACAG	CGCTACAGTC	AATGTCTATC	6900
GTAACAAAGA	CGGTAAACCA	GACTTGGACA	ACATCGTAGC	AACTAAAAAA	GTCACTATTA	6960
ACATAAACGG	TTTAATTTCT	AAAGAAACAG	TTCAAAAAGC	CGTTGCAGAC	AACGTTAAAG	7020
ACAGTATCGA	TGTTCCAGCA	GCCTACCTAG	AAAAAGCCAA	GGGTGAAGGT	CCATTACAG	7080
CAGGTGTCAA	CCATGTGATT	CCATACGAAC	TCTTCGCAGG	TGATGGTATG	TTGACTCGTC	7140
TCTTGCTCAA	GGCATCTGAC	AAGGCACCAT	GGTCAGATAA	CGGTGACGCT	AAAAACCCAG	7200
CCCTATCTCC	ACTAGGTGAA	AACGTGAAGA	CCAAAGGTCA	ATACTTCTAT	CAATTAGCCT	7260
TGGACGGAAA	TGTAGCTGGC	AAAGAAAAAC	AAGCGCTCAT	TGACCAGTTC	CGAGCAAACG	7320
GTACTCAAAC	TTACAGCGCT	ACAGTCAATG	TCTATGGTAA	CAAAGACGGT	AAACCAGACT	7380
TGGACAACAT	CGTAGCAACT	AAAAAAGTCA	CTATTAACAT	AAACGGTTTA	ATTTCTAAAG	7440
AAACAGTTCA	AAAAGCCGTT	GCAGACAACG	TTAAGGACAG	TATCGATGTT	CCAGCAGCCT	7500
ACCTAGAAAA	GGCCAAGGGT	GAAGGTCCAT	TCACAGCAGG	TGTCAACCAT	GTGATTCCAT	7560

317

ACGAACTCTT CGCAGGTGAT GGCATGTTGA CTCGTCTCTT GCTCAAGGCA TCTGACAAGG	7620
CACCATGGTC AGATAACGGC GACGCTAAAA ACCCAGCTCT ATCTCCACTA GGTGAAAAACG	7680
TGAAGACCAA AGGTCAATAC TTCTATCAAG TAGCCTTGGA CGGAAATGTA GCTGGCAAAG	7740
AAAAACAAGC GCTCATTGAC CAGTTCCGAG CAAACGGTAC TCAAACCTAC AGCGCTACAG	7800
TCAATGTCTA TGGTAACAAA GACGGTAAAC CAGACTTGGA CAACATCGTA GCAACTAAAA	7860
AAGTCACTAT TAAGATAAAT GTTAAAGAAA CATCAGACAC AGCAAATGGT TCATTATCAC	7920
CTTCTAACTC TGGTCTGGC GTGACTCCGA TGAATCACAA TCATGCTACA GGTACTACAG	7980
ATAGCATGCC TGCTGACACC ATGACAAGTT CTACCAACAC GATGGCAGGT GAAAACATGG	8040
CTGCTTCTGC TAACAAGATG TCTGATACGA TGATGTCAGA GGATAAAGCT ATGCTACCAA	8100
ATACTGGTGA GACTCAAACA TCAATGGCAA GTATTGGTTT CCTTGGGCTT GCGCTTGCA	8160
GTTTACTCGG TGGTCTAGGT TTGAAAAACA AAAAAGAAGA AAATAATCA GCTAAGGAAA	8220
TAAATGATGG ATAGTGGGCT GACTAAGATT AGTTTAAACA CTCAATCAGC AATCAGGACT	8280
TTCTTTCAAT AGCAGATTAA AATCATCGTA AAACAATAAA AATAGTGTTA TACTTAAAGC	8340
AGTATAGCAC TGTTTTTATC AAAGGAGAGA CAGATGGGAA AGACAATTTT ACTCGTTGAC	8400
GACGAGGTAG AAATCACAGA TATTCATCAG AGATACTTAA TTCAGGCAGG TTATCAGGTC	8460
TTGGTAGCCC ATGATGGACT GGAAGCGCTA GAGCTGTTCA AGAAAAAACC GATTGATTG	8520
ATTATCACAG ATGTCATGAT GCCTCGGATG GATGGTTATG ATTTAATCAG TGAGGTTCAA	8580
TACTTATCAC CAGAGCAGCC TTTCCTATTT ATTACTGCTA AGACCAGTGA ACAGGACAAC	8640
ATTTACGGCC TGAGCTTGGG AGCAGATGAT TTTATTGCTA AGCCTTTTAG CCCACGTGAG	8700
CTGGTTTTGC GTGTCCACAA TATTTTGCGC CGCCTTCATC GTGGGGGCGA AACAGAGCTG	8760
ATTTCCCTTG GCAATCTAAA AATGAATCAT AGTAGTCATG AAGTTCAAAT AGGAGAAGAA	8820
ATGCTGGATT TAACTGTTAA ATCATTTGAA TTGCTGTGGA TTTTAGCTAG TAATCCAGAG	8880
CGAGTTTTCT CCAAGACAGA CCTCTATGAA AAGATCTGGA AAGAAGACTA CGTGGATGAC	8940
ACCAATACCT TGAATGTGCA TATCCATGCT CTTGACAGG AGCTGGCAAA ATATAGTAGT	9000
GACCAAACCT CCACTATTAA GACAGTTTGG GGGTTGGGAT ATAAGATAGA GAAACCGAGA	9060
GGACAAACAT GAAACTAAAA AGTTATATTT TGGTTGGATA TATTATTTCA ACCCTCTTAA	9120
CCATTTTGGT TGTMTTTTGG GCTGTTCAAA AAATGCTGAT TGCGAAAGGC GAGATTTACT	9180
TTTTGCTTGG GATGACCATC GTTGCCAGCC TTGTCGGTGC TGGGATTAGT CTCTTTCTCC	9240
TATTGCCAGT CTTTACGTCC TTGGGCAAAC TCAAGGAGCA TGCCAAGCGG GTAGCGGCCA	9300

318

AGGATTTTCC	TTCAAATTGG	GAGGTTCAAG	GTCCTGTAGA	ATTTGAAJCAA	TTAGGGCAAA	9360
CTTTTAATGA	GATGTCCCAT	GATTTGCAGG	TAAGCTTTGA	TTCCTTGGA	GAAAGCGAAC	9420
GAGAAAAGGG	CTTGATGATT	CCCCAGTTGT	CGCATGATAT	TAAGACTCCT	ATCACTTCGA	9480
TCCAAGCGAC	GGTAGAAGGG	ATTTTGGATG	GGATTATCAA	GGAGTCGGAG	CAAGCTCATT	9540
ATCTAGCAAC	CATTGGACGC	CAGACGGAGA	GGCTCAATAA	ACTGGTTGAG	GAGTTGAATT	9600
TTTTGACCCT	AAACACAGCT	AGAAATCAGG	TGGAAGTAC	CAGTAAAGAC	AGTATTTTTT	9660
TGGACAAGCT	CTTAATTGAG	TGCATGAGTG	AATTTCAGTT	TTTGATTGAG	CAGGAGAGAA	9720
GAGATGTCCA	CTTGCAGGTA	ATCCCAGAGT	CTGCCCGGAT	TGAGGGAGAT	TATGCTAAGC	9780
TTTCTCGTAT	CTTGGTGAAT	CTGGTCGATA	ACGCTTTTAA	ATATTCTGCT	CCAGGAACCA	9840
AGCTGGAAGT	GGTGGCTAAG	CTGGAGAAGG	ACCAGCTTTC	AATCAGTGTG	ACCGATGAAG	9900
GGCAGGGTAT	TGCCCCAGAG	GATTTGGAAA	ATATTTTCAA	ACGCCTTTAT	CGTGTGCGAA	9960
CTTCGCGTAA	CATGAAGACA	GGTGGTCATG	GATTAGGACT	TGCGATTGCG	CGTGAATTGG	10020
CCCATCAATT	GGGTGGGGAA	ATCACAGTCA	GCAGCCAGTA	CGGTCTAGGA	AGTACCTTTA	10080
CCCTCGTTCT	CAACCTCTCT	GGTAGTGAAA	ATAAAGCCTA	AAACCCCTTT	ACAAATCCAG	10140
CTATTCATGG	TAGAATAGAT	TTTGTGTGAA	ATATCAGCAG	GAAAGCATGA	AGCTCGTCAA	10200
CAGGTGTCTT	ATGACAAGTA	ACCTTGGCTG	TTTAGGCGAA	GGGATCTGCG	ACGG	10254

## (2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9769 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCGGCGACTA	TCGATAACAC	TGACTTGGT	AGCCCCACAT	TTTGGACAAC	GCATCCTTTC	60
CCTCCTTATC	GTTTTCTTTT	CATTATACCA	TTTTTTAAGC	GATTCCTCAA	ACAATTCTTC	120
TTTTTGCTTG	ACAAGTTTTT	TGTTTTGTTG	TATTATTTAA	TTAAGACAAC	AAGGTAAAAG	180
AAAGGAGACT	AAGATGTCCT	GGACATTTGA	CAACAAAAAA	CCCATCTATT	TACAGATTAT	240
GGAGAAAATC	AAGCTTCAGA	TTGTTTCCCA	TACACTGGAA	CCCAATCAAC	AACTTCCAAC	300
CGTGAGGAGC	TAGCTAGCGA	GGCTGGTGTC	AATCCCAATA	CCATCCAAAG	AGCCTTATCA	360
GACCTTGAAC	GAGAAGGATT	TGTCTACAGC	AAGCGAACAA	CTGGACGATT	TGTGACTAAG	420
GATAAGGAGC	TAATCGCCCA	GTCACGCAAA	CAATTATCAG	AAGAAGAATT	GGAACACTTC	480

GTTCCTCCA TGACCCATTT TGGCTATGAA AAAGAAGAAC TACCAGGCGT AGTCAGTGAT	540
TATATTAAAG GAGTTTAAGC CTATGTCATT ACTAGTATTT GAAAATGTAT CCAAATCATA	600
TGGAGCAACA CCAGCCCTTG AAAATGTTTC TCTTGACATT CCAGCTGGAA AAATTGTCGG	660
CCTTCTTGGG CCAAACGGCT CAGGAAAAAC AACCTGATT AACTAATTA ATGGCCTCTT	720
ACAACCAGAT CAAGGACGTG TCCTCATCAA CGACATGGAC CCAAGCCCAG CAACCAAGGC	780
CGTTGTAGCT TATTTGCCTG ATACGACCTA TCTCAATGAG CAAATGAAGG TCAAAGAAGC	840
CCTAACCTAC TTCAAGACCT TCTATAAAGA TTGTCAGATC TTGAACGGCG CCATCATCTA	900
CTTGACAGACC TGGGCATTGA TGAAAATAGT CGTCTCAAGA AACTATCAAA AGGAAACAAA	960
GAAAAGGTTT AACTGATTTT GGTATGAGC CGTGATGCTC GTCTCTATGT TTTGGACGAA	1020
CCCATTGGTG GGGTGGATCC AGCAGCCCGT GCTTATATCC TCAATACCAT TATCAACAAC	1080
TACTCACCAA CTTCTACCGT TTTGATTTCCT ACCCACTTGA TTTCTGATAT CGAGCCAATC	1140
TTGGATGAAA TTGTCTTCCT AAAAGACGGA AAAGTCGTCC GTCAAGGAAA TGTAGATGAT	1200
ATTCGCTACG AGTCAGGTGA ATCCATTGAC CAACTCTTCC GTCAGaATTT AAGGCCTAAG	1260
CAAAGGAGAT TATTTATGTT TTGGAATTTA GTTCGCTACG AATTTAAAAA TGTTAACAAG	1320
TGGTATTTAG CCCTCTACGC AGCCGTGCTA GTCCTTTCTG CCCTCATCGG AATACAGACA	1380
CAAGGCTTTA AAAATCTACC TTACCAAGAA AGTCAGGCTA CTATGCTACT TTTCTAGCT	1440
ACAGTCTTTG GTGGCTTGAT GCTTACACTT GGGATTTCAA CCATTTTCTT GATTATTAAA	1500
CGCTTCAAAG GTAGTGCTA CGACCGACAA GGCTATCTGA CTTTGACCTT GCCAGTTTCT	1560
GAACACCATA TCATCACAGC CAAACTAATC GGTGCCTTTA TCTGGTCATT GATTAGCACC	1620
GCTGTATTGG CTCTAAGTGC TCTTATTATT CTGGCTTTAA CAGCTCCAGA ATGGATTCTT	1680
CTTTCTTATG TGATTACATT TGTAGAAACA CATCTCCCTC AGATCTTTCT TACAGGTATA	1740
TCCTTCCTAC TAAATACTAT TTCAGGAATC CTCTGCATCT ACCTGGCTAT TTCCATTGGA	1800
CAGCTTTTCA ATGAATACCG TACAGCACTC GCTGTTGCAG TCTACATTGG TATCCAAATC	1860
GTCATTGGAT TTATTGAACT TTTCTTCAAT CTTAGTTCTA ATTTCTATGT CAATTCACTG	1920
GTAGGACTCA ATGACCATTT CTATATGGGA GCAGGTATAG CCATTGTTGA AGAACTCATA	1980
TTCATAGCTA TCTTTTATCT CGGAACCTAC TACATCTTGA GAAATAAGGT TAATTTGCTT	2040
TAAATAATTT TTACCTAGAT ATGTAACATA CTCATAGAAC AAAAGAGACC AGGCAAAAAG	2100
TCTTTAAAAT TAGAAAACGC ATAGTATCAG GTGTTGAATA TGTAAGTCC CCAAAAAGTT	2160
AGATTTTTTC TGTCTAACTT TTGGGGCAG TTCATAAGAA CCTTGGTAAT ATGCGTTTTT	2220

320

TGTGAGCTGA CTTATTTCTT TTTACTATAT CGCAAAATGA AATAAGAACG GAACGATGGG	2280
ATTTTGGAAAT TCAAATCAAT TTATAAGAAT GTTTTAGAAG TAATATTATC CTATTCCAGA	2340
TTCACTTCAC TATACAATTG AGTTTTCAAG CAACCTGTTT ACATAATGTG TACATAATTA	2400
GGTTCGTGAT TCCACCCCTT TCACCTTTAA AAACCTCGCT TTCGCAAGGC TCTTCTATTT	2460
ATAAGATAAG GCACGTTTAA AGGTTTTCCA AATCCCTAAA TCATCCGTTT GAAGAACGAG	2520
ACTAGCATAC ATGCGTCCGA TAAATCCTGT TGCTACCACC GCAAAAATCA CTGTAATAGC	2580
AAGTGAAATC CATGCTTCTG CTCCCCCGC ATAGTCATTA ATCGTTCGAA ACGGCATAAA	2640
GAAGGTCGAA ATAAAGGGAA TATAAGAACC AATCTTCAAG AGGAGATTGT CACCAGCTGC	2700
ACCTAGAGCT GTCACCTCAA AAAAACCACC CATAATCAAA ATCATCAAAG GCGACAAGGC	2760
TTTCCCTGAG TCCTCAGGAC GAGAAACCAT AGATCCTAGG AAGGCTGCCA AGACTACGTA	2820
CATGAAAAGA CTGATCAAAA TAAAGAGCAA GGTATTCAGT GAGATAGCAT CTCCCAAGTG	2880
ATCCAAAATA CCAGACTGAG CCAAGAATGG CAAATCTTTA AAGAGCAAAA CGGCAGCCAG	2940
ACCACCTACA ACATAGATCC CAATATGCGT TAAATCACT AGAAACAGAG CCATCATCCG	3000
CGCATAGAAA TAGTGACTTG CCCTTATGCT AGAAAAAACC ACTTCCATAA TTTTGGTGCC	3060
TTTTTCACTG GCAACTTCCT GAGCTGTTAC ACCCGCATAG GTAATCAGAA TCATATAAAG	3120
AAAGAATCCT AAGGCACCTG CTGCAATTGT TTGAATAAAC TTTTATTTT CTTGGCTTC	3180
ATCAATCTTT TCTGTGAATT GAATTGTCTG CGCTAAGCCT TTTTCTGCT CTTGAGACAA	3240
GGAAGCAGTT GAACGATTAA GCTGATTTTG CAGTTCATTG AGTGACCTG TAACCTCAAA	3300
TTTAATTCCA TTTTCAAGCG ATGTTTCGCC ATGATAAACT GCCTTAGAA CACTATCTTC	3360
TTGATCAATG GTCAAATAAC CTTTAAATTT TTCTTCTTTA ATTGCTTCTT TGGCACTTGC	3420
TTCTGCTTTA TAGTCGAAGT TAACACCATT TACATTTTC AGTCCTTCTG CTACAGATGG	3480
CAGTGTGTC ACTACTGCCA CTTTATTATT TTTAGCCATA GAAGAACCCT GGAGATGCCC	3540
AATTCCTACA GAGATTCCTA AAAAGAGGAA CGGCGAAATC ACCATAAAGA AGAACTCCA	3600
TGACTCGACA TGTCGAAGAT AGGTTTCCTT GATTACAACC CACATATTTT TCATACTTCC	3660
ACTCCTGATT CTAGTTTAAA GATTTCATCG ATAGTTGGCG CTTGTTGGTC AAATGTTGCG	3720
ATATATTGAC CTTGAGTCAA GATTGAGAAG AGTTCCCTTC CAGCGCTCTC ATCTCCAAA	3780
ATCAATTTC AACTGCCTTG TTTGGTCAAG CTCACCTGTT TGACATGAGG AAGATTTTCC	3840
AATCTTCTCT TGCTTCGTTT ACTTGAAACA AAGAGACGGG TTTTCCCGTA TTGATTGCGG	3900
ACATCCTGAA CTGGTCCGTG CAAGACCACA CGGCCATCTC GGATCATCAG AATATCGTCA	3960
CAAAGTTCCT CAACATTGGT CATGACATGG TCAGAAAAGA TAATGGTTGT CCGCGCTCTT	4020

321

TTTCCTGAAA AATGACTTGT TTGAGCAATT CTGTATTAAC TGGGTCCAAT CCACTAAAAG	4080
GCTCATCCAA GATAATCAGG TCTGGTTCAT GAATCAGAGT AATAATGAGC TGAATCTTCT	4140
GCTGATTTCC TTTTGACAGA CTCTTGATTT TATCTGTCAG CTTTCCTTTC ACTTCCAACC	4200
TCTTCATCCA TTGAGGGAGT TTTTCTTTGA CTTCTTTGGC ATCCATGCCT TTTAGAGTCG	4260
CCAAGTAGCG AACTTGTTCA AGAACTGTCA ATTTAGGCAT GAGATGCCGT CTTCAGGCAG	4320
ATAACCAATC CGAGCATAGG TCTCCTGACG AATATCCTGA CCATCCAGAC CGATTTCTCC	4380
CTGATATTCT AGGAATTTCA AAATACTATG GAAAATCGTT GTTTTTCCAG CACCATTTTT	4440
TCCGACTAGT CCAAAAATAC GACCTGGTCG CGCTTGAAAG TCAATACCAA ACAAACCTTG	4500
CTTGGATCCA AAACCTTTCT CTAGACTTCT TACTTCTAGC ATCTTTCACC TCCGAAATTT	4560
CTTGCACTCA TTATACTCCT TTTTGATAGC CTTTACAATG TTTTTGTCC ATTTTGTAGAA	4620
GACTATTGCT GTGTAAAATA TGGCCTGGAG CACTTTTATA CTCAATGAAA ATCAAAGAGC	4680
AAACTAGGAA GCTAGCCGTA GACTGCTCAA AGTACAGCTT TGAGGTTGCA GATAAACTG	4740
ACGAAGTCgA CTCAAAACAC TGTTTTGAGG TTGTGGATAG AACTGACGAA kCrTAaCTAT	4800
ATCTACGGCA AGGCGAAcTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT TAGTGATAAA	4860
TCCATTATAC AGCAGCAAAC TTAATTTATA CCTTCCGCTC CTCAACTGTC TATTTTTAAT	4920
CCTGAATTGT TATTTGAGTA ACTCCTTTTT CCTCGTAAAG TTTTCTTCCT CTAAAACCTC	4980
TGGAAAAAGG CTAATAGTTT CAGACAACAT TTTTATAAGA AACAAGTTCA TCTGTCAATT	5040
CAAGAAGGAG TAATCCTTTA TCTACTAATG GACGGAACAG AATTCAACCG CTGTCCGAT	5100
ATGTTTTCTA AGGATTATAT AGTAAAATGA AATAAGAACA GGACAAATTG ATCAGGACAG	5160
TCAAACTGAT TTCTAACAAT GTTTTAGAAG TAGATGTATA CTATTCTAGT TTCAATCTGC	5220
TATATCTATT ATGCACACCC CTATAGGATC TAATGAAAAT CACAACAGGC TCATTCATAG	5280
ATGGTTACCT AAGCCTAAGG GAACTAAGAA AACGACTACC AAGGAAGTCG CATTCATCGA	5340
AAAGTAGATT AACAACTATC CTAaaaaATG CTTGAACTAC AAGTCCCCCA GAGAAGACTT	5400
CTGGATGACT AACTTGAAC TGAATTTAG CAATAATTAA TTCACTATCT AACTATATTT	5460
AGTAATTATT TCAGAACTGA TTAATATTAA AATTAATAA CAATTCAAAG GATTCACTACT	5520
AGCCATAAAT TACGTCCATC AGAGAGAGAC TCTTACTACT TTTAGATTTT AGTCTTTCTA	5580
GCTTCAGAAAT ACATCTAAAC TTTAGGGAAA ATGACTATTG GAAAGCGCGA ATGCCTCAAA	5640
ATTATCTCAG ATAAGCTATT CGAAACTTAG AATGCTTTTA AATTTATGGA ATTGCGATTA	5700
TTCGAAACCT AGAATGCATA TAACCTTTAG TTGACAGACC TATTCTAAGT CTCGAAGGGC	5760



322

TATTTACTTT CTATTCCTTA TCAAAAAAGA CTCATTCCCC CTTTCTCCTC CAAAATATGG	5820
TATAGTAGAA ATATACTATC TATGAGGAGT TTACATGTCA CAGGATAAAC AAATGAAAGC	5880
TGTTTCTCCC CTTCTGCAGC GAGTTATCAA TATCTCATCG ATTGTCCGGT GGGTTGGGAG	5940
TTTGATTTTC TGTATTTGGG CTTATCAGGC TGGGATTTTA CAATCCAAGG AAACCCCTCTC	6000
TGCCTTTATC CAGCAGGCAG GCATCTGGGG TCCACCTCTC TTTATCTTTT TACAGATTTT	6060
ACAGACTGTC GTCCCTATCA TCCAGGGGC CTTGACCTCG GTGGCTGGGG TCTTTATCTA	6120
CGGGCACATC ATCGGGACTA TCTACAATA TATCGGCATC GTGATTGGCT GTGCCATTAT	6180
CTTTTATCTA GTGCGCTAT ACGGAGCTGC CTTTGTCCAG TCTGTCGTCA GCAAGCGCAC	6240
CTACGACAAG TACATCGACT GGCTAGATAA GGGCAATCGT TTTGACCGCT TCTTTATTTT	6300
TATGATGATT TGGCCCATTA GCCCAGCTGA CTTTCTCTGT ATGCTGGCTG CCCTGACCAA	6360
GATGAGCTTC AAGCGCTACA TGACCATCAT CATTCTGACC AAACCCTTTA CCCTCGTGGT	6420
TTATACCTAC GGTCTGACCT ATATTATTGA CTTTTCTGG CAAATGCTTT GACACGTAAA	6480
AAATCCGTTT GGTTCCTCAA GTGGATTTTT AAAGCGTAGA TTAAGTATAG CTTGATACTA	6540
AATATACTTT GGTATGAAA TCATGCATAT TTTTCGATAG TGAGGCGAGG ACTTACCTAG	6600
CCTTCCGCC GTGATAGAAA CACCTGAAAT CTAATGGTTT CAGGTATTCG GAAACTTTGA	6660
GCCTAGTGTC TCAAAGTTTA GGTATGGAAT TTTGAAGAAA GTCGCTACCG TCCGTAATCA	6720
CTTAAGGAAA GGCTCAAAAA TATTGTTTTT AACCACAAAA TCCGTTTGGT TTCCCAAGCG	6780
GATTTTGTGC TTTATTTTGA AACTTCTTTT GCAAGAACAA AGTTCCCAAG TGTGGCAGAA	6840
CCATTTCTTG CGACTGCTGG CGTCACGATA TAGTCACGCA CATCTGGTAC TGGTAGGTAA	6900
CCATTAAGAA CAGATGTAAA TTTCTCACGG ACACGGTCCA GCATATGTTG TTGAGCCATG	6960
ACCCCTCCAC CAAAGACAAT CACGTCTGGG CGGAAAGTCA CTGTCGCATT AACCGCAGCT	7020
TGAGCGATAT AGTAGGCTTG AACATCCCAA ACAGGGTTGT TGAGTTCAAT AGTTTCCCCA	7080
CGTACACCTG TACGAGCTTC CAAACTTGA CCAGCTGCA7 AACCTTCTAG ACATCCCTTA	7140
TGGAAGGAC AAACACCCCT AAACCTTTTT TCAATATCCA TTGGGTGTCT AGCAACATAA	7200
TAATGACCCA TTTCAGGGTG ACCCACACCA CCGATAAACT CACCACGTTG GATGACGCCT	7260
GCACCGATAC CTGTACCGAT TGTGTAGTAA ACCAAGTTTT CGATACGACC ACCAGCATTG	7320
TTACGGGCAA CCATTTACCC GTAAGCAGAG CTGTTTACGT CTGTTGTGAA GTACATTGGC	7380
ACGTTTAGGG CGCCACGAAG GGCACCAAGC AAGTCTACAT TTGCCAGTT TGGTTTGGGA	7440
GTCGTCGTGA TAAAGCCATA AGTTTTGAG TTTTGTCAA TATCAATCGG CCCAAATGAA	7500
CCAACTGCAA GACCAGCAAG GTTATCGAAT TTTGAGAAGA ACTCAATGGT TTTATCGATT	7560

323

GTTCGATTG GAGTTGTTGT TGGAAATTGT GTTTTTTCTA CAACGTTAAA GTTTTCATCA	7620
CCGACAGCAC AGACAAACTT TGTACCGCCC GCTTCCAAGC TTCCATATAA TTTTGTCTATG	7680
ATAAACCTCT TGTTTTTATT TTCTTTATTA TAGCATACTT CGAAAGTCTA AATGTCTCTA	7740
TTTTTTAGAT TTCTCTCTGT AAATCTTACT ATCTAATAAA AACGAACAAA CATGTCATTT	7800
GTTCGTTTTC ACATTAGAGA GGATTGATTA GATTTTCACT TCGATCACAG CATCCCCCTT	7860
AGCAACTGAA CCTGTTGCCA CTGGAGCTAC TGAAGCGTAG TCACCTGTAT TTGTAACGAT	7920
AACCATTGTT GTATCATCAA GTCCAGCTGC AGCGATTTTG TTTGAGTCAA ATGTTCCAAG	7980
AACATCGCCA GCTTTCACCT TATTACCTTG AGCAACTTTT GTTTCAAAAC CGTCACCGTT	8040
CATAGATACA GTATCAATAC CAACATGAAT CAAAACCTCA GCACCATTTT TTGTTTTCAA	8100
ACCAAAAGCG TGCCCTGTTG GAAAGGCAAT TGAAACTCA GCATCAGCTG GTGCATAGAC	8160
CACGCCTTGG CTGGGTTTCA CAACGATACC TTGTCCATA GCTCCACTTG AGAAGACTGG	8220
GTCATTGACA TCAGCAAGAG CGACAACATC ACCGACGATA GGAGTTACAA GTGTTTCATT	8280
TTGAAGAGCT GCTGGCGCAA CTCTTCTTT TTCTTCAGCC ACTTCAGCTC GTTTTGCACC	8340
TGCAGTTGCG TCTACTTCAT CTTCGTAACC AAACATGTAA GTAAGAGCAA AACCAGGGC	8400
AAATGATACA GCTACCATAA GAAGGTATG TGAAGTTGT CCGTTACCAA CATAAAGCAT	8460
TGTACCAGG ATGATGGTGA TACCATTACC AGTACCAGCA AGTCCAAGGA TAGAAGCCAA	8520
TCCACCACCG ATTGCACCAG CAATCAATGA AAGGAAGAAT GGTTTACGGA AGCGCAAGTT	8580
CACCCCGAAG ATAGCAGGCT CTGTATACG TAGGAAGGCA GAAAGAGGAG CCGGGAAAGC	8640
AAGTGTTTTC AGTTTTGGAT TTTTGTTTT AACACCAACC GCAACAGTAG CAGCACCTTG	8700
AGCTGTCATA GCAGCTGTGA TGATAGCGTT GAATGGGTTA GCATGGTCAG CAGCAAGTAA	8760
TTGCACTTCA AGCAAGTTGA AGATGTGGTG CACACCTGAC ACGACGATCA ATTGGTGAAC	8820
CCCACCAATC AAGAAACCAC CAAGACCAAA TGGCATGCTA AGAATCGCTT TTGTAGCAAT	8880
AAGGATGTAG TTTTCAACAA CGTGGAACAC TGGTCCAATG ACAAAGAGTC CAAGGATAGA	8940
CATGACCAAA AGTGTCACGA ATGGTGTTAC CAAGAGGTCA ATGACATCTG GAACAACTTG	9000
CGGACAGCTT TTTCAAATTT AGCTCCGACA ACCCCGATGA TGAAGGCTGG AAGAACGGAA	9060
CCTTGCAAAAC CAACAACAGG GATGAAACCA AAGAAGTTCA TCGCTGTTAC TTCACCACCT	9120
TGAGCAACTG CCCAAGCGTT TGAAGTGAG CCAGAGACAA GCATCATACC AAGAACGATA	9180
CCAACGGCAG GATTTCCACC AAATACACGG AAGGTTGACC ACACAACCA ACCTGGCAAG	9240
ATGATGAAGG CTGTATCTGT CAAGATTTGT GTGTAAGTTG CAAAGTCACC TGAAGTGCC	9300

324

ATTTCAAGAG CGTTGAAAAG ACCACGCACA CCCATGAAGA GACCTGTCGC TACGATAACT	9360
GGGATGATTG GAACGAAAAC ATCACCACAAA GTACGGATAG CACGTTGGAA CCAGTTCCTT	9420
TGTTTAGCAA CTCTGCTTT CATGTCATCC TTAGATGATG TTGGTAATCC AAGTACAACA	9480
ACTTCATCGT ACATTTTGTT AACTGTACCT GTACCAAAGA TAATTTGGTA TTGCCCTGAG	9540
TTAAAGAAAG CACCTTGAAC TTTTCCAAG TTCTCAATCA CTCTTTTATT GATTTTCTCT	9600
TCATCTTTGA CCATGACACG TAGACGAGTC GCACAGTGGG CAACACTATT GACATTTTCA	9660
CGTCCGCCCA AGGCATCGAT GACTTTTMTT GCAATTTCTT GATTGTTTAT TTGCAAAAAT	9720
CTCCTTATAT AACATTTTGT TCTTGTTTGA AAGCGATTTT ATTCGCCGG	9769

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3149 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CGCTTGAGTG CTAATTCATA GTTCTATTGT ATCACTTGGT CAGAAATAAT CAAGAAAAAA	60
GTCTGACTTT CTCAAGATAA AAAGCCTGAG ACCAACTCAG ACTTTTAAAT TCTTAAATG	120
GCAATTCTTC CTCTCCAAG ACCAAATCTG CCAAATCTTG GCCTGCATTA TTTTCACGCA	180
TAGCACGTTG GGCACGACTT TCCAAGAGTT GGAATCCTGT GACAAGTACT TCGGTCACGT	240
AGTTCAATTG GCCATTTTTC TCAAAGCGAC GGGTACGCAA TTCTCCATCA ACGGAAATGA	300
GACTACCTTT GGTTCGTAC TTGCCAAAGT TTCTGCTAGT CTGCCCCATA GGACCATATT	360
GACAAAATCA GCTTCACGTT CACCGTTTGT GTCTTTGTAA CGACGGTTCA CAGCGATAGT	420
TGCTCGCGCT ACCGACTTGT CATTGTGGT TTTGTGCAAT TCTGGTGTAG ACGTTAAACG	480
TCCAATCAAG ATAACTTTAT TATACATATT TTCTTCTCC TACTTATCTA TTCGTAGGA	540
ATCAAAAAAA GTTACAGAAA TTTGTAACCT TTCGAGAAAA TTTTATTATT TTTATGAACC	600
ATGAAACCTG TCGCCTGTTG ATTGGCCATA ATGGTCATAT CTGTAATCTG AACACGACGA	660
GGTTGACTAG TCACATAGAC TACTGTATCT GCAATATCCT GAGCTTGCAA AGCTTCTATT	720
CCTTGGTAAA CGGACGCAGC TCGTTCTTTA TCACCATGAA AACGCACTGT AGAAAAATCT	780
GTTCGACAA TTCCAGGCTG AATGGTCGTC ACCTTGATAT CCGTTGCGAT GGTATCAATT	840
CGCAGTCCAT CTGAAAAGGT CTTAACTGCC GCCTTGGTGG CTGAGTAAAC AGCTGCACCA	900
GCATAGGCAT AAATTCCTGC GGTGACCCC ATATTGATAA TATGACCTTG ATTGGCTTTT	960

325

ACCATTGCTG GCAAGAAACA GCGAGTGA CTTCATCAAAC CTTTGACATT GGTATCCAAC	1020
ATGGTCAGCA TATCCAATC TTCATAGTCT TGATAGGGAG CTAAGCCAAG AGCCAGTCCT	1080
GCGTTATTGA CCAGGATGTC AATCTGACCT ATCGTTTCTA AAATATCAGA GCAGACAGTC	1140
TTTACCATTG TCATATCCGT GACATCTAGG AGAAAAGTCC AAATGTTTG ATTTGGAAAA	1200
GTTTCTGCAA ACTCCGCCTT AAGAGCTTCT AGTCTGTCTA TCCGTCGTCC TGTTAGAACG	1260
ACATCCTCAC CCTGCTCCAG ATAAGCACGC GCAATCGCTT CACCGATTCC TGATGTCGCT	1320
CCTGTAATCA CAACATTTT TGCCATCTTA TTCTCTCTA GCTGGTCTAT CAGATATTAA	1380
CAACTTCTTA GGCAGTCCAG TGTTCGCTG GGTGGAACGG TGTTCCGACA ACTTGGTCTT	1440
CTGATAATTC AAGCACCCCA CGTTTTGTG GAGCATTTGG CAGATGCAAT TCACGAGGAC	1500
TGCACATCAT ACCAAAACTC TTTTCAACAC GAAGTTCACC TGGGAAAATG AGATTCCCTT	1560
TTGGCATCAT AGCTCCAGGA AGCGCGACAA TGGTTTTCAA CCCCACACGC GCATTGGGAG	1620
CTCCTGCAAC GATTTGTACA GTCTTATCAC TTGCGACTGC AACTTGGCAG ATGTTGAGGT	1680
GGTCACTATC TGGATGGGCT ACCATCTCAA CAATTTCAAC TACAACAAA TTAGGTTCTT	1740
TATCATTAAC AATTTCTTCT GTAAAACCTT CCGCTGCAA CTCTTGGTTC AAACGAGCGA	1800
CTTGCTCATC TGTCAAAAAG ACTTGACCGC GCTCTGCAAT TTCAAATAAA CTTGAAACTT	1860
CGAAAAATTT CCAAGCCACT GTTTCCCAT TATCTTTGAG AAAAACACGG GCTACCTTGC	1920
CTTTGGCTC CACATCCAGT TTGGCATCTC CGCTATTTTT CACGATGACC ATAAGGACAT	1980
CACCGACATG TTCTTTATTA TATGTAAAA TCATTGTTTC CTTTTCTCC TATTTAGTC	2040
CTGCTAAAAA GTCATTGATT TGTGCTTGC TTTTACGGTC GCGATTGACA AAACGACCGA	2100
TTCTCTTGT CTTTTCTAGA ACAACAAGGC TAGGAATTCC GTAAACATCC CAGAGTTTGG	2160
CCAAATCCAT ATACTGATCT CGGTCCATTC GAATAAAGGT GAACTCTGGA TTGGTCTCCT	2220
CAATCTCTGG TAAGGCAGGA TAAATATAAC GACAATCGCT ACACCAGTCT GCCACAAAAA	2280
TGAAGACCTT CTGCCCCGCT TTTTCCACTA AAGATGCTAA TTCTTCTAAA CTTGCTGGCT	2340
GTATCATAAG ACTTCCTCCT CATAGACTAG GTCTTCATTT TCATAGACAA AGGTATAATG	2400
ACGGCCATCC TCAAAAATGA CGCCACCAAC CAAGCTCTCC AGACTGCTTT CGTAAACTTG	2460
AACATAAAGG GTCGCAATTT CCCCATGTC GGAAAAATGG TCTCGCACA TCTCTGTCAA	2520
CTCTTCTGA GTCTTCATGA GCTTACGGTC ATCTGCAACT TTTTCTGTAG CAAGAGCAAG	2580
GCTTCCGATA CCTAGCAGAG CCAAGCCTGC CATCCACATT TTTTGTAGCT TCATACCATT	2640
CATTTTAACA CAAAAAAGGC TTCAGGACAA ATGAGGAAGC AGCAGAAAAA CAAGTAAAAA	2700

326

GCCTCTTCCT TTAAGGAAAA GGACTTCTTA TACTCAATGA AAATCAAAGA CCAAAGTAGG	2760
AAGCTAGCCG CAGGCTGCTC AAAGCACTGC TTTGAGGTTG TAGATAGAAC TGACGAgTca	2820
CTCAAAACAC TGTTTTGAGG TTGTGGATGA AGCTGACGTG GTTTGAAGAG ATTTTCGAAG	2880
AGTATTATTC TTATTGCCAG GCACCTAAGT TGCCAACGTA GTAACATCA GGTGTGTAGG	2940
TATTGCGAGC ATCTTACCTG ATGAAGCCAG ATAATACTAC TTGCCATTGT CTTTGACCCA	3000
ATCATTCGCA ATCATGGAAC CAGAAGAACT TACATAATAC CATCTCCCT TGTCATAAAC	3060
CCAAGTACTG ACTTTCATGG TTCCTGAGCA ATTAAAGGCA AAAAACTGT CCAATAACAT	3120
TCGTTTTTTA AAAGCATTG ACACTACAT	3149

(2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10240 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

CCAAAAATTC AACCTTTAAG GGGAGTCCAG AGAGACTCAC AAGGTGTCAG ATAAAAGAAT	60
GGTGCAATTT TCTAGAGGAG ACTTTTGTAG TGTGCTCTCT TGTGTTGTAC GATTTTAACT	120
GAGGCCTTGC ACTAGCAAGG TCTTTTCTTT ATCTGGTCCC CTAAAAATTT AAGGAGGAAA	180
AGTTATGAAT CCCACATGTA AGAAGCGTTT GGGTGTCAAT CGGTTGGAAA CCATGAAGGT	240
GGTGCACAA GAGGAAATCG CGCCACAATC TTTGAATTAG TCCTAGAAGG AGAAATGGTT	300
GAAGCCATGC GAGCAGCCCA ATTTCTTCAT CTGCGTGTAC CGSACGATGC CCATCTCTTA	360
CGTCGTCCTA TTTCAATTTT GTCTATTGAC AAGGCAAACA AGCAGTGTCA CCTCATTTAT	420
CGGATTGACG GAGCTGGGAC TGCAATTTTT TCAACCTTAA GTCAGGGAGA CACTCTTGAT	480
GTGATGGGGC CTCAGGGAAA TGGTTTGTAC TTGTCTGACC TTGATGAGCA GAATCAGGT	540
CTCCTTGTTG GTGGTGGGAT TGGTGTCCA CCCTTGCTTG AGGTGGCCAA GGAATTGCAT	600
GAACGTGGAG TGAAAGTAGT GACAGTCCTC GGTTTGTCTA ATAAGGATGC TGTATTTTG	660
AAAACGGAAT TGGCTCAGTA TGGTCAGGTC TTTGTAACGA CAGATGATGG TTCTTATGGC	720
ATCAAGGGAA ATGTTTCCGT TGTTATCAAT GATTAGACA GTCAGTTTGA TGCTGTTTAC	780
TCGTGTGGGG CTCCAGGAAT GATGAAGTAT ATCAATCAAA CCTTTGATGA TCACCCAAGA	840
GCCTATTTAT CTCTGGAATC TCGTATGGCT TGTGGGATGG GAGCTTGCTA TGCTGTGTT	900
CTAAAAGTAC CAGAAAACGA GACGGTCAGC CAACGCGTCT GTGAAGATGG TCCTGTTTTC	960

327

CGCACAGGAA CAGTTGTATT ATAAGGAGAA AATTATGACT ACAAATCGAT TACAAGTTTC	1020
TCTACCTGGT TTGGATTGCA AAAATCCGAT TATTCCAGCA TCAGGCTGTT TTGGCTTTGG	1080
ACAAGAGTAT GCCAAGTACT ATGATTTAGA CCTTTTAGGT TCTATTATGA TCAAGGCGAC	1140
AACCCTTGAA CCACGTTTTG GGAATCCAAC TCCAAGAGTG GCAGACACGC CTGCTGGTAT	1200
GCTCAATGCA ATTGGCTTGC AAAATCCTGG TTTAGAGGTT GTTTTGGCTG AAAAGCTACC	1260
TTGGCTGGAA AGAGAATATC CAAATCTTCC TATTATTGCC AATGTAGCTG GTTTTCAAA	1320
ACAAGAGTAT GCAGCTGTTT CTCATGGGAT TTCCAAGGCA ACTAATGTAA AAGCTATCGA	1380
GCTCAATATT TCTTGTCCTCA ATGTTGACCA CTGTAATCAT GGACTTTTGA TTGGTCAAGA	1440
TCCAGATTTG GCTTATGATG TGGTGAAAGC AGCTGTGGAA GCCTCAGAAG TGCCAGTTTA	1500
TGTCAAATTA ACCCCGAGTG TGACCGATAT CGTTACTGTC GCAAAAGCTG CAGAAGATGC	1560
GGGAGCAAGT GGCTTGACCA TGATCAATAC TCTGCTTGGG ATGCCCTTTG ACCTCAAAAC	1620
TAGAAAACCA ATCTTGGCCA ATGGAACAGG TGGAATGTCT GGTCCAGCAG TCTTCCAGT	1680
AGCCCTCAAA CTCATCCGCC AAGTTGCCCA AACAACAGAC CTGCCTATCA TTGGAATGGG	1740
AGGAGTGGAT TCGGCTGAAG CTGCCCTAGA AATGTATCTG GCTGGGGCAT CTGCTATCGG	1800
AGTTGGAACA GCTAACTTTA CCAATCCTTA TGCCTGCCCT GACATCATCG AAAATTTACC	1860
AAAAGTCATG GATAAATACG GTATTAGCAG TCTGGAAGAA CTCCGTCAGG AAGTAAAGA	1920
GTCTCTGAGG TAAACTGCAA TCAATCTGTT CTTGATTTTT TATTAGTTTG TAATATGAAT	1980
TTAGGAGAAT TTTGGTACAA TAAAATAAAT AAGAACAGAG GAAGAAGCTT AATGAACAAA	2040
GTAAGATTTA TTTTTTTAGC TCTGCTATTT TTCTTAGCTA GTCCAGAGGG TGCAATGGCT	2100
AGTGATGGTA CTTGGCAAGG AAAACAGTAT CTGAAAGAAG ATGGCAGTCA AGCAGCAAAT	2160
GAGTGGGTTT TTGATACTCA TTATCAATCT TGGTTCTATA TAAAAGCAGA TGCTAACTAT	2220
GCTGAAAATG AATGGCTAAA GCAAGGTGAC GACTATTTTT ACCTCAAATC TGGTGGCTAT	2280
ATGGCCAAAT CAGAATGGGT AGAAGACAAG GGAGCCTTTT ATTATCTTGA CCAAGATGGA	2340
AAGATGAAAA GAAATGCTTG GGTAGGAACT TCCTATGTTG GTGCAACAGG TGCCAAAGTA	2400
ATAGAAGACT GGGTCTATGA TTCTCAATAC GATGCTTGGT TTTATATCAA AGCAGATGGA	2460
CAGCACGCAG AGAAAGAATG GCTCCAAATT AAAGGGAAGG ACTATTATTT CAAATCCGGT	2520
GTTTATCTAC TGACAAGTCA GTGGATTAAT CAAGCTTATG TGAATGCTAG TGGTGCCAAA	2580
GTACAGCAAG GTTGGCTTTT TGACAAACAA TACCAATCTT GGTTTTACAT CAAAGAAAAT	2640
GGAAACTATG CTGATAAAGA ATGGATTTTC GAGAATGGTC ACTATTATTA TCTAAATCC	2700

328

GGTGGyTACA TGGCAGCCAA TGAATGGATT TGGGATAAGG AATCTTGGTT TTATCTCAAA	2760
TyTGATGGGA AAATrGCTGA AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC	2820
TACTTCAAAT CCGGTGGTTA CATGACAGCC AATGAATGGA TTTGGGATAA GGAATCTTGG	2880
TTTTACCTCA AATCTGATGG GAAAATAGCT GAAAAAGAAT GGGTCTACGA TTCTCATAGT	2940
CAAGCTTGGT ACTACTTCAA ATCTGGTGGC TACATGGCGA AAAATGAGAC AGTAGATGGT	3000
TATCAGCTTG GAAGCGATGG TAAATGGCTT GGAGGAAAAA CTACAAATGA AAATGCTGCT	3060
TACTATCAAG TAGTGCCTGT TACAGCCAAT GTTTATGATT CAGATGGTGA AAAGCTTTCC	3120
TATATATCGC AAGCTAGTGT CGTATGGCTA GATAAGGATA GAAAAAGTGA TGACAAGCGC	3180
TTGGCTATTA CTATTTCTGG TTTGTCAGGC TATATGAAAA CAGAAGATTT ACAAGCGCTA	3240
GATGCTAGTA AGGACTTTAT CCCTTATTAT GAGAGTGATG GCCACCGTTT TTATCACTAT	3300
GTGGCTCAGA ATGCTAGTAT CCCAGTAGCT TCTCATCTTT CTGATATGGA AGTAGGCAAG	3360
AAATATTATT CGGCAGATGG CCTGCATTTT GATGGTTTTA AGCTTGAGAA TCCCTTCCTT	3420
TTCAAAGATT TAACAGAGGC TACAAACTAC AGTGCTGAAG AATTGGATAA GGTATTTAGT	3480
TTGCTAAACA TTAACAATAG CCTTTTGGAG AACAAGGGCG CTACTTTTAA GGAAGCCGAA	3540
GAACATTACC ATATCAATGC TCTTTATCTC CTTGCCATA GTGCCCTAGA AAGTAAGTGG	3600
GGAAGAAGTA AAATTGCCAA AGATAAGAAT AATTTCCTTG GCATTACAGC CTATGATACG	3660
ACCCCTTACC TTTCTGCTAA GACATTTGAT GATGTGGATA AGGGAATTTT AGGTGCAACC	3720
AAGTGGATTA AGGAAAATTA TATCGATAGG GGAAGAACTT TCCTTGGAAG CAAGGCTTCT	3780
GGTATCAATG TGGAATATGC TTCAGACCCT TATTGGGGCG AAAAAATTGC TACTGTGATG	3840
ATGAAAATCA ATGAGAAGCT AGGTGGCAAA GATTAGTACT ATAAGTGAAT ATGATTTGAG	3900
TGAATAGTAA GTTAAAAATC CTGATTTCAA GTAAAATCAG GATTTTTTCA TGGATGCAAT	3960
TTTTTTGGAG TCTGGTGTGA CGCGGAGGGT CTTTTGTCTT GTGTAAGTGA CAAAGCCGGG	4020
TTTTCCACCA GTTGGTTTAT TGAGTTTTTT GACTTCAATC ATATCTACCT GCACCAGATT	4080
CGACAGGCGC CCTTGAGAGA AGTAGGCAGC TAACTCTGCT GCGTCTGTCT TGAATGCAATC	4140
AGATGGGTCA AGATTTCTTG AGATGACAAC ATGGCTTCCA GGAATGTCTT TAGCATGGAA	4200
CCAAAGTTCC TCCTTGCGGG CCATTTTAAA GGTCAATTCC TCATTTTGAA GATTGTTTCG	4260
TCCGACATAG ATGATGGTTT TGCCATCGCT TGCTAGATAT TGTTCTAGTT TTTTCCGTTT	4320
CTGGATTTTC TCCCGTTGTC TTCTGCGGAT AAAACCTGTT TGAATCAATT CTTACCGGAT	4380
TTCAGCGATT TCTTCCAGTC CAGCTTGGTT GAGGACGGTT TCTACACTTT CCAGATAGAC	4440
AATAGTGGCT TTGGTTTCTT CAATCAAATC AGTCAAGTAT TTGACAGCTT CTTTGAGTTT	4500

329

CTGATACCGT TTAAATAGC GTTGGGCATT CTGGTTGGGA GTCAGAGCCT TATCAAGCGC	4560
AATCATGATA GGTGGTTGG TATAGTAGTT GTCTAGGATA ACCTGGTCTT GGTGTTAGG	4620
CACTTGGTGG AGGAAGGTTG TCAGCAATTC TCCTTTTGA CGAAATTCTT CAGCGTTGTC	4680
TGTCGCCAGT AACTCTTTT CCTGTTTTT SAGTTTGTGT CGGTTTTCTT GAAGTTCATT	4740
TTCAACACGA CGAATCAGTT CACTGGCCTG CTGTTTGACG CGGTCGCGCT CAGCCTTATC	4800
CTTATAGTAG GTGTCCAACA AATCAGAAAG ATTTGCAAAA GGCTCTCCCA CCTGATTTGC	4860
AAAAGGAACT GGACTGAAGG AAGTCTCAGT CAAGCATGGC TTGGTTTCTT GATTGAAAAA	4920
ATTTGCGAAA GCGGAAAGTT TTCTACTAAC CAGTATCCTT TCCAATTCAT TTGCCGTATC	4980
GCGTCCCGA CCTTGAAAGA GGCTTTGAAG ATTTTTTGCT GTTACTTCTT GGGTTTGCAG	5040
GATTTCAAAG AGCTTTTCAT CCTTGATAGT AAAAGGATTG AGAGATTTTG TACTTGGCGG	5100
AGCGATATAG GTCGATCCTG GAAGTAAGGT GCGGTAGCTA TTTTGTGAAA AGCCGACGTG	5160
TTTGATAACT TCGAGGATTT TATGACTGCT TTTATCGACC AGTAGAATAT TACTGTGTTT	5220
CCCCATAATT TCGATAATCA AGGTAGCCTG GATATGGTCT CCAATCTCGT TTTTATTGGA	5280
AACTGTAATT TCCACAATAC GGTCATTTTC CACTTGCTCA ATCGACTCAA TCAGGGCCCC	5340
CTGCAAATAC TTTCTCAAAA CCATGATAAA GGTAGAAGGT TGAGCTGGAT TTTCAAAAGT	5400
CGTTTGGGTC AGCTGAATGC GTCCAAAAAC TCGATGGGCA GAAAGGAGCA GCGCATGGCT	5460
TTGGCGATTG CTGCGGATTT GCAAGACCAA CTCTTGTTCA AAAGGCTGAT TGATTTTCTG	5520
GATGCGACCA TTTCTAATT CGCTTCGCAA TTCTCAACT ATGTGGTGTA AAAAAATCC	5580
GTCAAATGAC ATCGTTCTCT CTTGTGATT GTATTCCATA GTATTATATC AAAAAGGTAG	5640
AATAAAATCA TGGAAATGTG GTATAATAAA GCCAAGTAAA GAGAAACGAG AAGCACATGT	5700
ATATTGAAAT GGTAGATGAA ACTGGTCAAG TTTCAAAGA AATGTTGCAA CAAACCCAAG	5760
AAATTTTGA ATTTGCAGCC CAAAAATTAG GAAAAGAAGA CAAGGAGATG GCAGTCACTT	5820
TTGTGACCAA TGAGCGTAGT CATGAACTTA ATCTGGAGTA CCGTAACACC GACCGTCCGA	5880
CAGATGTCAT CAGCCTTGAG TATAAACCGA AATTGGAAAT TGCCTTTGAC GAAGAGGATT	5940
TGCTTGAAAA TTCAGAATTG GCAGAGATGA TGTCTGAGTT TGATGCCTAT ATTGGGGAAT	6000
TGTTTCATCTC TATCGATAAG GCTCATGAGC AGGCCGAAGA ATATGGTCAC AGCTTTGAGC	6060
GTGAGATGGG CTTCTTGCCA GTACACGGCT TTTTACATAT TAACGGCTAT GATCACTACA	6120
CTCCGGAAGA AGAAGCGGAG ATGTTCCGTT TACAAGAAGA AATTTTGACA GCCTATGGAC	6180
TCACAAGACA ATAAACGAAA ATGGAAAAAT CGTGACTTGA TATCCAGTTT AGAATTTGCT	6240



330

TTGACAGGTA TTTTACTGC TATCAAGGAA GAACGCAATA TGGAAACA CGCAGTGACG	5300
GCTCTAGTGG TCATCCTTGC AGGTTTTGTT TTTCAGGTGT CACGAATCGA ATGGCTCTTT	6360
CTCCTATTGA GTATTTTCTT GGTAGTAGCC TTGAGATTA TCAACTCTGC TATTGAAAAT	6420
GTGGTGGATT TGGCCAGTCA CTATCACTTT TCCATGCTGG CTAAAAATGC CAAGGATATG	6480
GCGGCCGGCG CGGTATTAGT GGTTCCTCTT TTCGCAGCCT TAACAGGCGC ATTGATTTTT	6540
CTCCCACGAA TCTGGGATTT ATTATTTTAA ACAGTAAGAG GAAATTATGA CTTTAAATC	6600
AGGCTTTGTA GCCATTTTAG GACGTCCCAA TGTGGGAAG TCAACCTTTT TAAATCACGT	6660
TATGGGGCAA AAGATTGCCA TCATGAGTGA CAAGGCGCAG ACAACGCGCA ATAAATCAT	6720
GGGAATTTAC ACGACTGATA AGGAGCAAAT TGTCTTTATC GACACACCAG GGATTCACAA	6780
GCCTAAACA GCTCTCGGAG ATTTTCATGGT TGAGTCTGCC TACAGTACCC TTCGCGAAGT	6840
GGACACTGTT CTTTTCATGG TGCCTGCTGA TGAAGCGCGT GGTAAGGGGG ACSATATGAT	6900
TATCGAGCGT CTCAAGGCTG CCAAGGTTC TGTGATTTTG GTGGTGAATA AAATCGATAA	6960
GGTCCATCCA GACCAGCTCT TGTCTCAGAT TGATGACTTC CGTAATCAAA TGGACTTTAA	7020
GGAAATTGTT CCAATCTCAG CCTTCAGGG AAATAACGTG TCTCGTCTAG TCGATATTTT	7080
GAGTGAAAAT CTGGATGAAG GTTCCAATA TTTCCCGTCT GATCAAATCA CAGACCATCC	7140
AGAACGTTTC TTGGTTTCAG AAATGGTTCG CGAGAAAGTC TTGCACCTAA CTCGTGAAGA	7200
GATCCCGCAT TCTGTAGCAG TAGTTGTTGA CTCTATGAAA CGAGACGAAG AGACAGACAA	7260
GGTTCACATC CGTGCAACCA TCATGGTCTGA GCGCGATAGC CAAAAAGGGA TTATCATCGG	7320
TAAAGGTGGC GCTATGCTTA AGAAAATCGG TAGCATGGCC CGTCGTGATA TCGAACTCAT	7380
GCTAGGAGAC AAGGTCTTCC TAGAAACCTG GGTCAAGGTC AAGAAAAACT GCGCGGATAA	7440
AAAGCTAGAT TTGGCTGACT TTGGCTATAA TGAAAGAGAA TACTAAGTAG AGGTAGGCTC	7500
ATGCCTGCTT CTTGTTTTTA CAGAAGGAGG ACTTATGCCT GAATTACCTG AGGTTGAAAC	7560
CGTTTGTCGT GGCTTAGAAA AATTGATTAT AGGAAAGAAG ATTCGAGTA TAGAAATTCG	7620
CTACCCCAAG ATGATTAAGA CGGATTTGGA AGAGTTTCAA AGGGAATTGC CTAGTCAGAT	7680
TATCGAGTCA ATGGGACGTC GTGGAAAATA TTTGCTTTTT TATCTGACAG ACAAGGTCTT	7740
GATTTCCCAT TTGCGGATGG AGGGCAAGTA TTTTACTAT CCAGACCAAG GACCTGAACG	7800
CAAGCATGCC CATGTTTTCT TTCATTTTGA AGATGGTGGC ACGCTTGTTT ATGAGGATGT	7860
TCGCAAGTTT GGAACCATGG AACTCTTGGT GCCTGACCTT TTAGACGTCT ACTTTATTC	7920
TAAAAAATTA GGTCTGAAC CAAGCGAACA AGACTTTGAT TTACAGGTCT TCAATCTGC	7980
CCTTGCCAAG TCCAAAAGC CTATCAAATC CCATCTCCTA GACCAGACCT TGGTAGCTGG	8040

331

ACTTGGCAAT ATCTATGTGG ATGAGGTTCT CTGGCGAGCT CAGGTTTCATC CAGCTAGACC	9100
TTCCCAGACT TTGACAGCAG AAGAAGCGAC TGCCATTCAT GACCAGACCA TTGCTGTTTT	8160
GGGCCAGGCT GTTGAAAAAG GTGGCTCCAC CATTCCGACT TATACCAATG CCTTTGGGGA	8220
AGATGGAAGC ATGCAGGACT TTCATCAGGT CTATGATAAG ACTGGTCAAG AATGTGTACG	8280
CTGTGGTACC ATCATTGAGA AAATTCAACT AGGCGGACGT GGAACCCACT TTTGTCCAAA	8340
CTGTCAAAGG AGGGACTGAT GGGAAAAATC ATCGGAATCA CTGGGGGAAT TGCCTCTGGT	8400
AAGTCAACTG TGACAAATTT TCTAAGACAG CAAGGCTTTC AAGTAGTGA TGCCGACGCA	8460
GTCGTCCACC AACTACAGAA ACCTGGTGGT CGTCTGTTTG AGGCTCTAGT ACAGCACTTT	8520
GGGCAAGAAA TCATTCTTGA AAACGGAGAA CTCAATCGCC CTCTCCTAGC TAGTCTCATC	8580
TTTTCAAATC CTGATGAACG AGAATGGTCT AAGCAAATTC AAGGGGAGAT TATCCGTGAG	8640
GAACTGGCTA CTTTGAGAGA ACAGTTGGCT CAGACAGAAG AGATTTTCTT CATGGATATT	8700
CCCCTACTTT TTGAGCAGGA CTACAGCGAT TGGTTTGCTG AGACTTGGTT GGTCTATGTG	8760
GACCGAGATG CCCAAGTGA ACGCTTAATG AAAAGGGACC AGTTGTCCAA AGATGAAGCT	8820
GAGTCTCGTC TGGCAGCCCA GTGGCCTTTA GAAAAAAGA AAGATTTGGC CAGCCAGGTT	8880
CTTGATAATA ATGGCAATCA GAACCAGCTT CTTAATCAAG TGCATATCCT TCTTGAGGGA	8940
GGTAGGCAAG ATGACAGAGA TTAAC TGGA GGATAATCTG CGCATTGCCT GGT TTGGTAA	9000
TTTTCTGACA GGAGCCAGTA TTTCTTTGGT TGTACCTTTT ATGCCCATCT TCSTGGAAAA	9060
TCTAGGTGTA GGGAGTCACC AAGTCGCTTT TTATGCAGGC TTAGCAATTT CTCTCTCTGC	9120
TATTTCCGCG GCGCTCTTTT CTCCTATTTG GGGTATTCTT GCTGACAAAT ACGGCCGAAA	9180
ACCCATGATG ATTCGGGCAG GTCTTGCTAT CACTATCACT ATGGGAGGCT TGGCCTTTGT	9240
CCCAAATATC TATTGGTTAA TCTTTCTTCG TTTACTAAAC GGTGTATTTG CAGGTTTGT	9300
TCCTAATGCA ACGGCACTGA TAGCCAGTCA GGTTCCAAAG GAGAAATCAG GCTCTGCCTT	9360
AGGTACTTTG TCTACAGGCG TAGTTGCAGG TACTCTAACT GGTCCCTTTA TTGGTGGCTT	9420
TATCGCAGAA TTATTTGGCA TTCGTACAGT TTTCTTACTG GTTGGTAGTT TTCTATTTTT	9480
AGCTGCTATT TTGACTATTT GCTTTATCAA GGAAGATTTT CAACCAGTAG CCAAGGAAAA	9540
GGCTATTCCA ACAAAGGAAT TATTTACCTC GGTAAATAT CCCTATCTTT TGCTCAATCT	9600
CTTTTAAACC AGTTTGTGCA TCCAAATTTT AGCTCAATCG ATTGGCCCTA TTTTGGCTCT	9660
TTATGTACGC GACTTAGGGC AGACAGAGAA TCTTCTTTTT GTCTCTGGTT TGATTGTGTC	9720
CAGTATGGGC TTTTCCAGCA TGATGAGTGC AGGAGTCATG GGCAAGCTAG GTGACAAGGT	9780

332

GGGCAATCAT CGTCTCTTGG TTGTCGCCCA GTTTTATTCA GTCATCATCT ATCTCCTCTG	9840
TGCCAATGCC TCTAGCCCCC TTCAACTAGG ACTCTATCGT TTCCTCTTTG GATTGGGAAC	9900
CGGTGCCTTG ATTCCCGGGG TTAATGCCCT ACTCAGCAAA ATGACTCCCA AAGCCGGCAT	9960
TTGAGGGTC TTTGCCTTCA ATCAGGTATT CTTTATCTG GGAGGTGTTG TTGGTCCCAT	10020
GGCAGGTTCT GCAGTAGCAG GTCAATTTGG CTACCATGCT GTCTTTTATG CGACAAGCCT	10080
TTGTGTTGCC TTTAGTTGTC TCTTTAACCT GATTCAATTT CGAACATTAT TAAAAGTAAA	10140
GGAAATCTAG TGCGAGTAAA AATCAATCTC AAATGCTCCT CTTGTGGCAG TATCAATTAC	10200
CTAACCGTA AAAATTCAAA AACCCATCCA GACagATTGA	10240

(2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13206 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CGCTTTATCG TGGACGTGGT CAAGCCGAGA ATTTTCATCA GGAGATGAAG GAGGGATTTT	60
TTGGCGATAA AACGGATAGT TCAACCTTAA TCAAAAACGA AGTTCGTATG ATGATGAGCT	120
GTATCGCCTA CAATCTCTAT CTTTTTCTCA AACATCTAGC TGGAGGTGAC TTCCAAACTT	180
TAACAATCAA ACGCTTCCGC CATCTTTTTC TTCACGTGGT GGGAAAATGT GTTCGAACAG	240
GACCGAAGCA GCTCCTCAAA TTGTCTAGTC TCTATGCCTA TTCCGAATTG TTTTCAGCAC	300
TTTATTCTAG GATTAGAAAA GTCAACCTGA ATCTTCCTGT TCCTTATGAA CCACCTAGAA	360
GAAAAGCGTC GTTAATGATG CATTAAAGAA CAGTCGAGAT GAAAAAATCG TGTGACGCAC	420
CAAGGGAGGA GTCTGCCCTT TTGAGGAAAT CTAGCGAGGA AAAACGATAC TGGAACAGCA	480
GAAAGTAAAA CTGACCTCAT GAGGAGGAAG AAAGTGGCTC ATGAGGTCAG GGGTTTGTAA	540
AGTTACATCT AGTTGAGAGA GGTATGAATG ATTTGGGATT AATCATTCTT TGTTTTAAAT	600
CAGGAGAATA GTAACGATTT TTTCCTTTTT TGACGAACTC TATTCCGTAA CGATCAATCA	660
ATTTAATCAT GTACCTAATA TTAGAATTGT TTATCCCAAA TTTATTTGAA AGCTTCTCTA	720
AGCTATATCC TTGTTTTCTA AGTTCATAGA TCTGAACTTT ATCATCATAA GTTAGTTTCA	780
TAATAAAAAAC ACCCCAAAAG TTAGATTTTT TCTGTCTAAC TTTTGGGGGG CAGTTCATTC	840
AACACCTGAT ACTATGCGTT TTTCTTATTT GAAATACTTT TTACTCAACC TCTTTATACT	900
CAATGAAAAT CAAAGTGCAA ACTAGAAAGC TAGCCTCAGG CTGCTCAAAA CAGTGTTTTG	960

333

AGGTTGCAGA TGGAAGCTGA CGTGGTTTGA AGAGATTTTC GAAGAGTATT ACTTAATCTT	1020
CTTGATACTT TGACTAAGAA TAAATCCTAC AATCATCCCT ACCATATTTT GCATAAAATT	1080
CGGTAGAATT TCTGGGAGGG CTGCTGCCCA GCCATTCATC AAAGCAGAAC CCAAGGCGTA	1140
GCCTCCTACC ATGGCAATAG TTGCTAAAAT AAGGCCTAAC CACTGACTTT TTCCTTTAAA	1200
TCCTGCGAAA AATCCCTGCA AGCCATGGTT GACCAAGCTA AAGAACATCC ACTGAGGGTA	1260
GCCTGATAAG AGGTCAATCA AGAACTTGC TAGTCTCCG ACTACCGCTC CTTCACGACT	1320
ACCAAAGTAA AAGGCCGCAA AGAAGACACC AGCATCTAAA AGAGTTAGAA TTCCTGTAGG	1380
TGTTGGGATT TTTAAGAAAT AACCTAGAAC CACAGAAAGG GCGGTTAATA GGGATACAAG	1440
GGCGATTTTA GTTGTTTTGG TTTGCTTCAT ATTGTCTTAC TCCATACTGA TCTGCTTGTC	1500
CAATAGCAGC ATAAACGAAA GCCTTAGAGC TTTCTACTGC TGGCAAAAGT TTATCACCTT	1560
TAACCAGGTG ACTGGCAATG CTAGAGSCAA AGGTACAACS TGCACCAGCA TTTTGGCCTT	1620
GGATAACTGG ATTTTCTAGG ATAGTAAAGG TCTGTCCATC ATAAAAGACA TCCACAGCCT	1680
TGTCCTGACT AAGACGATTG CCTCCCTTGA TAATGACTGE GGCCTCCTA AATCATGCAA	1740
TTTCTGCGCT GCAGTTTTC A TGTCTTCCAA GGTTTTAATT TCCTGACCGG ATAATAATTC	1800
TGCTTCTGGG AGATTAGGCG TAATCACACT GACATAAGGG AAAAAGCGAA TCAACTCTTG	1860
GCAGAGCTCA CTGACAGCTA CATCATGCGT TTCCTTGCG AGCAAGACAG GATCCAACAC	1920
CACAGGTACT CCTGGGCGTT GTTTGATAAA GTCCAAGGCC TTCTCAGCCA CGCTGACAGT	1980
AGGGAGAAGA CCAATCTTAA TTCCCCCAA TTCCACATCA CGCAAGCTAT CTAATTCATG	2040
TTGAAAAATG GTATCATCAG TTGGAAAGAC TTCAAATCCT TTTTCTGTCA AGGCTGTCAA	2100
ATAAGTCACT GCTACAAAAC CATGCAAGCC GTTCAAGGT TAGCTAGCCA ATCAGGCTGA	2160
CAGTCCACCA CCACTAAAAA TATCATTTCC AGAAAGTGCT AAAATACGAT TATTCTTCAT	2220
AACGAATCTC CTTTAAATAC AAACCATTTG GTGCTGCAGT GGGACCTGCA AGTTGCCTGT	2280
CCTTCTTCTC CAAGATGAGA TCAATCTGCT CTA CTGGCAT GCGGTTGTTA CCGATTTTGA	2340
GAAGAGTCCC CACCATATTG CGAATCTGTT TATACAAGAA ACCATTTCTT GAAAAGGTAA	2400
AGGTCAAAAA TTGCTCTGTC TCATCGACTA TTAAACTAGC TTCTGTGATG GTGCGAACCT	2460
TATCTCTAC ACTAGTCCCA GAGGCTGTAA AACCGGTAAA ATCATGGGTT CCTCTAGCT	2520
TTTTGATTGC AATCTGCATT CGTTCACAT CGAGTGGGTA GGGAAAGTGG GTGGCATAGT	2580
GACGGCGCAT CGGATTTTGG GGACGTCTC TATCCACAGT AAACATAG GTCTTGCTAT	2640
GCTTGGCATA ACGGCAATGA AAATCATCTG CCACAAGCTC AATCGAAATC ACATCAATAT	2700

334

CTTCAGGAGA CTGGGTATCC AAGGCAAAAC GGAGTTTCTC CTCATCCATC TGATAAGGCA	2760
GGTCAAAATG AATCACCTGT CCCAGGGCAT GAACCCCACT ATCTGTCCTA CCAGCACCGT	2820
GAACAGTAAT GGCTTGCCCT TTATTTAATC TGGTCAAGGT TTTTTC AATT TCTTCCTGAA	2880
CGCTACGCGC ATGAGGCTGG CGCTGAAAGC CAGCAAAGGC ATAACCATCA TAGGAAATAG	2940
TTGCTTTATA TCTCGTCATA GCCTCTATTT TATCAAGAAA TTAGTCTGTA AACAAAGGACC	3000
TAAAACAAAT ATTGTATGGG TATAAAAATC TCATACTCTT CGAAAATCTC TTCAAACCAC	3060
GTCAGTTTCC ATCTGCAACC TCAACACACT ATTTTGAGCA ACCTGCGGCT AGCTTTCTAT	3120
AGTAGATTGA AATAAGATAT GAACAACTCT ATTAGGAAAG TCAAATTAAT TTCTAGAAAT	3180
ATTTTAGCAG CTACAGCGTA CTATTCCAAA CTCAATCAAC TATAGTTTGC TCTTTGATTT	3240
TCATTGAGTA TCAAAAGAAA AACTTAGGAA TCAATCCTAA GCTCTCTTCT GAAGTAGGTA	3300
CATGACAAAG ATAGAGATTA CAATCAACCA ACCTCCTAAG ATACTAAAGA CCAACATCCC	3360
ATTGTGAGTT AGTAAGCCAA TTGCACCTAG AACGAATGGG GTCGTAAAGG CTCCGAAACT	3420
ACAGCCTAAT ACAGCAAATG AAGTTGCTTG ATTGAGGACT TTAGCTGGAA TTCGTT CAGA	3480
GACAAGTTGA AAGACCGTCG TCAAGACTAC ACTATAGGCA AATCCAGCCA GAACACTTCC	3540
TGCTACTACC ACCCACAAGG ATGAAGACAA GGCAATCACG ATTTGCCCCA AGCCAAAGGT	3600
AATACCAGAC CAGAGGAGCA GTTTCTCTTT AAAGATAGAA ATCAAGAAAG AAAA ACTCAC	3660
CCCAGCCACA ATCCCGATCA ACTGCATGAT ACTAAGAACA AA ACTAGATA ACTGGGCATC	3720
CCCCAATCCT CTTTCCACCA TCAA ACTTGG AATACGGATG GTAATAGCTG TATTGGTACA	3780
AACTACA ACT GCCGCTTCGA TAGCTAAGGT AAAAATCAAG CCTTTCATTT CTCGAGTTAA	3840
ACGACTTGCT TCCTTCGCTC TTTTCTTGAC TTTCTCTTT GATTTTCCAT AAGGGACAAA	3900
GAGCAGATAA AGGGGCAGCA CCAAAAATCC AGCACTATAG GCTAGAAAGA TAGCTGTCCA	3960
ACCAAAGGCC AACAACTGAC CGACGGCCAA GGTAATGAGA GAAGCTCCAA CGACCTCTGC	4020
AGAAGCGCGT AGCCCTAACA TCTGAATTCG CCTTTTCTCT TGGTAGCGTT CACTGATAAT	4080
AGAAATGGCC TTGGCATTGA TCATCCCAAG ACCCAAACCA AAGAGAAGCC GTGTTCCAAA	4140
GACAAAGGGA TAGGCTTGGT ACCAGAAGGG AGCTGTACCG CTCAATGATA AAATCAGCAA	4200
GCCCCAACTA ATCTGTAAGC GCTCAGGAAA TATTTTTTCT AAGAAACCAT TTAGCAGTAA	4260
CATCATCATG ATTCCAAAGG AAGGCAAGCT CACCAAGAGC TCAATTGTG CCTTAGAATA	4320
ACCCTGATAA TAGTCAAACA TGCTGGTAG GGCACCTGAA ATGGAAAAGG AGGTAATCAA	4380
AACGAGGGAG AGAGCCAAAA TGCTGGCCCG TTCTAAAAAT TGTTTCATGA AATCTCTTTC	4440
TATATTCTC TTAATCTTCT ACTTTTTTGA TAGTTATCAA ATAAGCAAGA AAAGAAGAAG	4500

335

CCTCATTGGT TTGTAGACTC CTTCTTAAAT TCGAAAATGA ATCCCTTGTA TCTTATACTC	4560
AATGAAAATC AAAGAGCAAA CTAGGAAGCT AGCCGCAGGT TGTTCAAAAC AGTGTTTGA	4620
GGTTGCAGAT GGAAACTGAC GTGGTTTGAA GACATTTTCG AAGAGTATTA GGATGACTTT	4680
CTCTTGATT GCTTGATAAA GTAGAAAATA AATCCTGCTA CCATATAGGC AACAAAGATA	4740
ATCAGACACC ACTTAAACAC AACATTCCAA CCCTTGTTCA CATTCAAAAA GAAGTAAGGG	4800
AAAGGATTAT CCTTGGCATT TGGAAATATG AGTTTTAGAA CCAAGCCATT AAAAAGAGCA	4860
AACATCATAT ACAGAAAGGG TAAATGGTC CACACTGCTG GATCCCAAAT CTTGTATTGA	4920
CCCTGTTTGT CAAAAACAG GGTATCCGCT AAAAACCAGA TGGGAACGAT ATAGTGGCAA	4980
AGGAAATTTT CTAGGTGATA GAAATTAGTC GCAATGGGCG CCAAGAGGAA ATGGTAAATC	5040
ACACAGGTAA TCATGATACT CATGGTGACC CCACCTTTTA AGCGCAAGAG ACTTGGCCTT	5100
TGCCAATTTT CACCTACAGC GCTCATAACC TTTAGAAGAT AAAGGGTAAA AATAGTTACC	5160
AAGAGGTTGG ACAGAACCGT GTAATAGAGA AGCATCCCAA AACCACCATG CTTAGTAATT	5220
TCAAGATAAA CTCCCGTAAA AGCCGCTAGA AACAAGAAGA TACGGCTATA AAATACAAGT	5280
TTATAGTGTT TTGACATGCT TAAATCTTCC TCACAAACTC TGATTTAAGT TTCATGGCAC	5340
CAAAACCATC AATCTTACAG TCGATATTGT GGTGCGCTTC TACGATGCGG ATATTTTCA	5400
CGCGCGTCCC TTGTTTCAAA TCTTTGGCG CACCTTTTAC TTTCAAGTCC TTGATGAGAG	5460
TTACTGTATC ACCATCAGCC AATTTATTTT CGTTGGCATC GATAGCGACA AGACCTTCTT	5520
CTACTTCTGC AACTTCAGCA GGATTCCACT CATGAGCACA CTCTGGGCAA ACCAGTAGGG	5580
CACCGTCTTC GTAGACATAC TCTGAGTTAC ATTTTGGACA ATTTGGTAAA TTGTTGATGG	5640
TTTCTCCTTA TCATCATTCG CTATCTTTT AAAATCAAAA TTTCTGCAAC AGCAACTATT	5700
ATACCCTAAA ATCAGCATTG TGACAAATTG AGAAAAAAC CGATATCAAT CTATCGGCTT	5760
TTCTACATTG ACATTCTTTT TTCAGCTTCT GCTTTGATTG TTTCAACTAC TTCTTGAATG	5820
TTCAAACCAG TTGTATCAAG GTAGACAGCA TCCTCTGCTT GTTTGAGAGG AGAAGTCTCA	5880
CGATGACTAT CCTTGTAAGT ACGCGCAGCA ATTCCTTTT TTAGGGTTTC AAGGTCTGTT	5940
TCAATTCCTT TGGCAATATT TTCCTTGTA CGACGCTCTG CTCTCTCATC AACAGAAGCT	6000
ACTAGGAAAA TTTTCAATTC TGCTTGTTGGC AATACAACAG TTCCAATATC GCGACCATCC	6060
ATGACAATCC CGCCTTGCTG GGCAATTTCT TGTGGAGAG AAACCAGTTT CTCACGCACT	6120
TGAGGAATTG CTGCAATAGC AGAAACATGA TTGGTCACTT CATTTTCACG GATAGGATGG	6180
GTAAATATCCA CATCTCCTAC AAAAACAAGC TGGTCTCCAG TTTCTGAACG TCCAAAGCTG	6240

336

ATTGGATGCT GGTCCAACAA GGCTAGAAGG GCTTCGACTT CTTCAACTCC TAATTGGTTC	6300
TTAAGAGCCA TATAGGTCGC TGCACGATAC ATAGCTCCTG TATCAAGGTA GGTGAATCCA	6360
AAATCCTTAG CAATAATCTT TGCACCGTA CTCTTACCGC TGGAAAGCAGG ACCATCAATA	6420
GCAATTTGAA TTGTTTTCAT ATCGGCTCCT ATTTTATTTT TATAACATCA CCTGGATTAG	6480
CAAACCAAGA TCCTGTAGCC ATGTGCCCAG GATTCAAGGC CTCTAACTGA GCAATGGAGA	6540
TTCTGTCAGC AGCGGCAATA GCTGCTTCCC CTTCTCCTGC GAGAACTTTA ATCGTTCCCTT	6600
CAGGATTAGC AGCTTCTTCT GAACTACTAG AAGTAGATTG TGGCTCTGAA CTCTGCTCAG	6660
GCTGAGAACT ACTTGAAGAT GAGATTGTGA CTACACTGGC ATCAGAATCA TGAAAGCCTT	6720
TTAAGGCTGC TGTGCGATTA CTCCCCCCCG ATGATAGATA GATGAGAACC ATGACCATCA	6780
CCACCACAAT TACAAAGAAA ATACTAGCTA GGATCGTCAA AATACGATTA GCCATCCTAT	6840
CAGCCCCCTCC GTGGTTTTCGA TGCCGACGCT CTGCTCTTGA TTCTTCTTGA TCATAGATAT	6900
CTTCTTGCCA CGGTTCTTTT GCCATACCTT ACTCCTTGTT TTTTTTTACT TTTCTTATTA	6960
CAATATAAAT ATGAACATGA AAATCACACT TATACCTGAA CGATGTATCG CCTGTGGGCT	7020
TTGCCAAACT TATCTGATT TATTTGATTA CCACGATAAT GGAATCCTGC GTTTTTACGA	7080
TGACCCTGAC CAACTGGAAA AAGAAATTC TCCTAGTCAG GATATCTTAG AGGCTGTAA	7140
AAATTGCCCA ACTCGCGCCC TGATTGGAAA CCAGGAAGCC TAAATCAATG GCGATAATCC	7200
ACTCCCTCTA GTTTAGCACA TTTCCATGTA AAATTATAGT CTTTTCACCTT TATTTTTTTC	7260
TGTAAATCA GGAAGGTCAC TTTTTCTTTT GATAAGATAA AGTGGTCTTT TTTTAGTCTC	7320
TAAATAAATC TTAATGATAT ACTTGCCGAG AATCCCAATG GTCAAGAGTT GAATGCCTCC	7380
AAGAAAGAGA ATAACAGCCA TCAGAGAGGT CCAACCAGAT GTCCGATTGC CCAAAATGAG	7440
GGTCCGAACC ACAACAAAAA AGGTGATCAG CAGAGAAAGA AAACAAGATA GGAGACCAGC	7500
TACAAAGGCT ATAATCAAGG GAAAATCTGA AAAATTAATA ATCCCTTCAA TGGAGTAGAA	7560
AAAGAGTTGC CTAAACTCC AACTTGTCTT GCCAGCCTGC CTTTCGACAT TTGGATAGTC	7620
CAAATAGTAG GTTTTGAAAC CCACCCAGGC GAAGAGCCCC TTGAAAAAC GATTGGACTC	7680
GGTCAAGCTT AAAATGGCAT CGACTACAGA CCTTCTCATC ATACGAAAAT CACGGACACC	7740
CGACGGCAGA GCTACTGGGC TGATTTTTTG CATGAGGCGA TAAAAGAGAA CAGCACAGAA	7800
ACTGCGAAAG AAGGGTTCTC CCTCCCGACT AGTTCTCCGT GTCCCAACGC AGTCCAAGTC	7860
TACATTTTGT TCTAATACAT TTTTCATCTC AAACAACATA CTAGGAGGAT CTTGGAGGTC	7920
TGCATCCATC ACCACCACCA AATCTCCTGT CGCATATTGC AAGCCTGCAT AAAGGGCTGC	7980
TTCTTTGCCA AAATTTCGAG AGAAAGAAAT ATAATGGACT GCCGGATTTT GCTCCCGATA	8040

337

GGCCTTTAAG AGTTCCAAGG TCCCATCACT TGATCCATCA TCGACAAAGA CATACTCGAT	8100
TTCTGTTTCC AAATCTGGAA GTAAAGCTTC CAGAGCCTGA TAAAAAAGAG GAAGTACTTC	8150
CTCTTCGTTT AAACAAGGGA CGATGATTGA AATCATCATC TTAGTCTTCA AATCCATTG	8220
GATGCTTGCT TTGCCAACGC CATGCGTCTT CACACATTTG GGTGATGTCT AGTTCTGCTT	8280
CCCAACCGAG TTCTGCTTTA GCTTTTGCCG GGTCTGAGTA GCAGGCAGCG ATATCACCTG	8340
GGCGACGTC TACGATGCGG TAAGGAATAG GACGGCCAC CGCTTTTCC ATGTTTGGGA	8400
TAATTTCAAG AACTGAGTAA CCTTTACCAG TTCCAAGGTT ATAAACGTTT AGTCCTGAAC	8460
CTTTTGGAT TTTTTCAAA GCTGCAACGT GACCCTTAGC CAAATCGACA ACGTGGATAT	8520
AGTCACGAAC ACCTGTTCCA TCTCCGTAT CGTAATCGTC TCCAAACACT TGCACCTGCT	8580
CTAATTTTCC AACGGCTACT TGAGTCACAT ATGGCAAGAG ATTGTTTGGGA ATACCGTTTG	8640
GATTTTCTCC CAAATCACCA CTCTCATGGG CTCCGATTGG GTTAAAGTAA CGAAGCAAGA	8700
CAACATTCCA TTCTGAGTCT GCTTTGTAAA TATCAGTCAA AATTTCTCT AGCATGAGCT	8760
TAGTACGACC GTATGGGTTG GTCACGTAAA GTGGGAAATC TTCCAAGATG GGCACGTGTG	8820
GCGGATCCCC GTAAACTGTC GCAGAAGAAC TGAAGATGAT GTTTTACAG TTGTTTCTT	8880
CCATGGCTTT CAAAAGGCTG ACAGTTCAG CGATATTGTT GTCATAGTAG GCAAGAGGGA	8940
TACGTGTTGA TTCGCCAACA GCCTTCAAAC CAGCAAAGTG AATGACACCA GTCGGTTCTT	9000
CCTGCTTGAA AATATCTCTG AGGGTATCTG TGTCACGAAT ATCTGCCTCA TAGAAAGGAA	9060
TCTCAACTCC TGTGATTCTT TCAACAACCT TAAACTCTT ACGATTGCTA TTGACAAGAT	9120
TATCCACCAC AACAACTTGA TGACCTGCTT GGATCAATTC AATAACAGTG TGGGTTCCAA	9180
TAAACCGGC ACCACGATT ACCAAAATCT TTTCTTGCAT TTTTCTCTT CCAATCTCAG	9240
ATTATTTTTT CTTATTTTAC CATTTTGTAC AGGGAATGTC ATTTGCCATC CTAAACTACC	9300
TGATAAAAT TCAGTAAAT GCTTATACTC TTCGAAAATC CAATTCAAAC TACGTCAACG	9360
TCGCCTTGCC ATGGGTATGG TTA CTGACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT	9420
GTMTGAGCT GACTTCGTCA GTCTATCCA CAACCTCAA GCAGTGCTTT GAGTAACCCG	9480
CGGCTAGTTT CCTAGTTTGT TCTTTGATTT TTATTGAGTA TTATTCGCTT TTTACTCGTT	9540
TGACATAGTT TTCAATTGGG TAATTTAGAG GGTCCAAGGT CAACTCCTTG TCTTGGATCA	9600
GTTGGGCTAG ATGGTAACCA ATGATAGGAC CAGTTGTGAG GCCTGATGAA CCTAGTCCAC	9660
TGGTGCATA GACACCAGTT AAGTCAGGCA CCTGCCAAA GAAAGGAGAG AAATCACTGG	9720
TGTAGGCACG GATTCCAACA CGCTCAGATT TTGAAGTAGC TTCAGCCAAA ATCAGATAGT	9780



338

GAGTCAAGGT GGCCTCCTCC ATTTGTTGGA GCAAGGTTTC ATCTACCGTC AAATCAAATC	9840
CCATGTCATT TTCGTGGGTA GCGCCTAAGG ATAATTTCCC ACCTGCAAAG GGAATCAAAT	9900
CCCCTCCCC TTCTGGCATG ACAACAGGGT AATCTTCCAT GTCTTGGGCA AGCTGATAAT	9960
CTCGTAGTTG TCCTTTTGA GGACGGACAT CCACCTCATA ACCTAAAGGC TCTAACATGT	10020
CCCCCAACCA AGCTCCCGTC GCCAAAATAA CCTGCTCAAA CTCCTCTTCA CCAATCTGGT	10080
AGCCTGATGC TAACGGTGTG AGAGTCACTT TTTCTTTGAC CAGCTTGACA TGACTGACTT	10140
CCAGCAAACG AGTCACTAAA AGTTGGCCAT CTA CTCTCGC TCCACCAGAA GCATAGAGCA	10200
GGCGGTCAAA TCCTTGCAAA CCAGGGAATA ATTCATTAGC TGAGGCTTGG TTGAGAATGG	10260
CTAATTGCCC TATCAAGGGA GATTCTTCTC TGCGCTGGAG GGCCAGTTGA TAAAGTTCTT	10320
CCAAATTGGA TTCATCCTTT TTCAAGAGAA AGACTCCCGA ACGCTGGTAA AAGTCGATT	10380
CTTGTCCTGA TTTCTCTAAA TCAGCTAATA AATCCACATA AAAATCAGCC CCCAAGCGCG	10440
CCATCTTGTA CCAGGCTTTA TTACGGCGTT TGGAAAACCA AGGACTGATA ATTCCTGCTG	10500
CGGCCTTGGT GGCTTGACCT TGCTCATGGT CAAAAACGGT CACCTCTAGG TCACTTTCTC	10560
TCGAGAGGTA GTAGGCAGCT GTTGCTCCCA CAATTCTGTC TCCAATAATG GCAACTTTTT	10620
TCATTGTCTT CACTTTCTAA CTAGATATGA TGGAAAGGAT TGGTTGATGC CTGACTAGGC	10680
AAGATATCAA TAGACCACCC CTTATCTTCC TTCCATTGAC TAAGAAGTGC TGCGATTTTT	10740
TCTACAAAAA TCACTTCGAT ATAGTGACCT GGGTCCAATG CAAGCAACCC ATCAGATAGC	10800
ATATCCTGAG CAGTATGGTA GTAGATATCA CCAGTGATAT AGACATCTGC CCCCTTTGCC	10860
AAAGCATCCT TATAGAAAGA CTGCCCCGCTT CCACCACAAA TTGCTACTCT TGAAATAGGC	10920
TTCTGCAAAT CATCCTCTTG ATAATGCACC ATTGGAAGGC TATCTAGGTC AAAGACTTGC	10980
TTGACCTGTT GGGCCAATTC CAAAATGTC TGAGGCTGAA TATCCCAAT ACGTCCAATT	11040
CCACGTTCTG GACCTGTTTC CTGCAGATAA GTCGTCTCCT CGATTCTTAG CATCTGACAA	11100
AACCAGTCAT TGAGCCCATT TTCAACGATA TCAATATTGG TATGGCTGAC ATAAACTGCC	11160
ATATCATGCT TAATCAGGTC GATGTAAATC TGATTTTGCG GACGGCTGGC AAGCAAGTCC	11220
TTGATAGGAC GAAAGATAGG CGCGTGCTTG ACGATAATCA AGTCCACACC CTTTTCAATG	11280
GCCTCTGCCA CTGTCTCTTC ACGAATATCG AGGGCAACCA TGACCCCTTG GATACCCCTG	11340
TCTAAAGTGC CAATTTGCAG ACCACGGCTG TCTCCCTCCA TAGAAAATTC CTGAGGGCAA	11400
AAGGCTTCAT AAGCTTGAT CACTTCACTT GCTAACATGG AGCACCTCCT TGATAGCTTG	11460
AATCTTATCT ACTAGAACTT GACGTTCTTC CAGATTTTTT TCTGGGATTT GTCCGAGGGC	11520
GAACCTAGC TTCTCAGCTT CTTTTGCCA TTTTGGACA AATACTGGAC TGACTTCTTT	11580

339

GGACAAGAAG GGACCAAAGC GAACATCACT GGCTGATAGC TTCATTTGTC CTGCTTCCAC	11640
CACCAAAATC TCATAAAACT TTCCAGCTTC TTCTAAGATG CTTTCTGCTA CAATCTGGAA	11700
TCCATGATCC TGTAGCCAGA TACGCAAGTC GTCTTCACGA TTATTGGGCT GGAGGATCAA	11760
ACGCTCTACA TTAGCTAACT TCCCCAAACC TTCTTCTAAA ATCCTAGCAA TCAAACGACC	11820
ACCCATGCCA GCAATGGTAA TGACAGACAC TTGGTCAGTC TCTTCAAAAG CTGCCAAGCC	11880
ATTGGCTAAA CGGACTTGGA TTTTCTCCTT TAGGCCGTGA GCCTCAACAT TTTAACC GC	11940
AGACTGATAG GGACCTTCCA CCACCTCACC TGCAATAGCG CTTTGTATTG GCCTCTCTC	12000
AACCAACTCG ATAGGCAGAT AAGCATGGTC ACTTCCCACA TCTAGTAAAA TAGCCCCCTG	12060
TGACACAAAG GAAGCTACCA ATTCTAATCT CTTTGAAATC ATCTTCTCTC ACTTTCCAAA	12120
ACTCTATTAC CTCTTATTAT ACCACATTTT AATCTTCAAC TTTCCAGTAA TATAAGCACC	12180
TCTGGCGAAA GAAGTTTCAA TGTCTTAAAG TAATAAGTGA ATCCAATTGA AAGATTTTAA	12240
ACAAATTTGCA AAAATGTCAA AAAATAAAAA ATAAACAGTT TATTCAGAAA ATTCTTGACA	12300
TATAAAAACA CATGGTAGAA TATAATTAGA AAGTTAGAAA AAATAAAAGT TTGACTAAAA	12360
TTTGTATTTG AAGGTGGTGT TCAGATAAGA AATTAGTCA GACGAACCAC GAATTTGCTC	12420
TATGCTTTCT GGAATTTATC ATAACAGGAG GATACAGTCA TGGAAACAAAC ATTGTTTGAA	12480
TTAGAACTAC TTCCAGAGGA AGATATCATT GTCACAGGTC TCCCTAAGTA TTGTTCTTTT	12540
ACTTGTTTAA TTACAGGTCG CTAGTTATAT TTTATATAAA ATAAGTAGCT TTACTTACGG	12600
AATAGGCTAG TGCTGTGTCT CTAGCCTATT TTAATAATTA GGAGTTTGTT ATGGATTTAT	12660
TAGAGAAAGA ATGTTTAAAA TGTGATAAAA ATTTCCAACA GGGTGATATT TGGAATTACT	12720
ATTATTTATC AGATAAGATG CCTGCACAAG GGTGGAAAAT ACACATAAGC TCCCAAATAA	12780
AAGACGCTGT AAATATTTT AAGATTGTGT ATAACTATC CCAACTAAAT AATTGTAGCT	12840
TTAAAGTTGT TAAAAATTA GAGGAATTAA AAAAAATTAA TTCCCTAGG GAAATGAGCC	12900
CTACTGCTAA CAAATTTATA ACTCTATATC CTAAGTCAGA ATCTGAAGCT AAGACTATGA	12960
TTTGTAATCT TACGAATAGA CTGTCAGAAT TTAAGGCTCC AAAAATACTA TCTGACTATC	13020
AATGTGGAAT GCATTTCTCCA GTTCATTATA GATATGGGGC TTTTAAAAA AAACAAGCTT	13080
ATGATGAAAA AAATAAAAAA GTCATCTATT TATTGCTAGA TGAAAAAAGG AAGAACTATG	13140
TAGAAGATAA GAGACAAAAT TTCCCTAGTC TTCCTAGCTG GAAAATGGAT TTATTTTCAG	13200
AAGAAG	13206

(2) INFORMATION FOR SEQ ID NO: 34:

340

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13104 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

```

CCGGATCCAG CGAAAAATAT GCTCTTTGAT GCTGTAAGTG GTCAAAAAGA TGCTAAAACA      60
GCTGCTAACG ATGCTGTAAC ATTGATCAAA GAAACAATCA AACAAAAATT TGGTGAATAA      120
AAAATTTGTT CAAGGGGGGT GGAAATCAAA TCCCCCTTTG AATTTATCAA TAGAGACACA      180
AATAATTTAG CTTTCTTATA AAAAAGTAGT ATCCTATGAA AGGAGTTAAT ATGGAAAAGC      240
AACAACTAG TAAAGCAGCC CTGCTGTCTA TCATTCCTGG GTTAGGACAG ATTTACAATA      300
AACAAAAAGC CAAAGGTTTT ATCTTCCTTG GTGTAACCAT CGTATTTGTC CTTTACTTCC      360
TAGCACTTGC AACCCCTGAA TTGAGCAACC TCATCACTCT TGGTGACAAA CCAGGTCGTG      420
ATAATTCCTT CTTTATGCTG ATTCGTGGTG CCTTCCATCT AATCTTTGTA ATCGTTTATG      480
TACTCTTTTA TTTCTCAAAT ATCAAAGATG CACATACGAT TGCAAAACGC ATTAACAATG      540
GAATTCAGT TCCACGCACA CTCAAAGACA TGATCAAAGG GATTTATGAA AATGGCTTCC      600
CTTACCTCTT GATCATTCCA TCTTATGTTG CCATGACCTT CGCGATTATC TTCCCAGTTA      660
TCGTAACCTT GATGATCGCC TTTACCAACT ACGACTTCCA ACACTTGCCA CCAAACAAGT      720
TGTTGGACTG GGTGGGTTTG ACCAACTTTA CAAACATTG GAGCTTGAGT ACCTTCCGTT      780
CTGCCTTTGG TTCTGTTCTT TCTTGGACTA TCATTTGGGC TTTGGCAGCT TCTACTTTAC      840
AAATCGTAAT TGCTATCTTC ACAGCTATCA TTGCCAACCA ACCATTTATC AAAGGAAAAC      900
GTATCTTTGG TGTTATTTTC CTTCTTCCTT GGGCTGTCCC AGCCTTCATC ACTATCTTGA      960
CATTCTCAAA CATGTTTAAC GATAGTGTCC GTGCTATCAA CACTCAAGTA TTGCCAATCT     1020
TGGCTAAATT CCTTCCTTTC CTTGATGGAG CTCTTATTCC TTGGAAAACA GACCCAAC/T     1080
GGACTAAGAT TGCCTTGATT ATGATGCAAG GTTGGCTCGG ATTCCCATAC ATCTACGTTT     1140
TGACCTTGGG TATCTTGCAA TCTATTCCTA ACGACCTTTA CGAAGCAGCT TATATTGACG     1200
GTGCCAACGC TTGGCAAAAA TTCCGCAACA TCACTTTCCC AATGATTTTG GCTGTTGCGG     1260
CACCTACTTT GATTAGCCAA TACACCTTCA ACTTTAACAA CTTCTCTATC ATGTACCTCT     1320
TCAATGGTGG AGGACCTGGT AGTGTGGGAG GTGGAGCTGG TTCAACCGAT ATCTTGATCT     1380
CATGGATCTA CCGTTTGACA ACAGGTACAT CTCCTCAATA CTCAATGGCG GCAGCTGTTA     1440
CCTTGATTAT CTCTATCATT GTCATCTCAA TCTCTATGAT CGCATTCAAG AACTACACG     1500

```

341

CATTTGATAT GGAGGACGTC TAAGATGAAT AACTCAATTA AACTCAAACG TAGACTGACT	1560
CAAAGCCTTA CTTACCTTTA CCTGATTGGT CTATCAATTG TAATTATCTA TCCACTGTTG	1620
ATTACCATTA TGTCAGCCTT TAAAGCAGGT AACGTCTCAG CCTTTAAACT AGATACTAAT	1680
ATCGACCTCA ATTTTGATAA CTTTAAAGGC CTCTTCACTG AAACCTTGTA CGGTACTTGG	1740
TACCTCAACA CTTTGATTAT CGCCTTAATT ACCATGGCTG TTCAAACAAG TATCATCGTA	1800
CTTGCTGGTT ATGCTTACAG CCGTTACAAC TTCTTGGCTC GTAAACAAAG TTTGGTCTTC	1860
TTCTTGATCA TCCAAATGGT GCCAACTATG GCCGCTTTGA CAGCCTTCTT CGTTATGGCG	1920
CTTATGTTGA ACGCCCTTAA CCACAACCTG TCCTCATCT TCCTCTACGT TGGTGGTGGT	1980
ATCCCGATGA ATGCTTGGCT CATGAAAGGC TACTTCGATA CAGTGCCAAT GTCTTTAGAC	2040
GAATCTGCAA AACTAGACGG TGCAGGACAC TTCCGCCGCT TCTGGCAAAT TGTTCACCA	2100
CTTGTTGCGC CAATGGTTGC CGTACAAGCT CTCTGGGCTC TCATGGGACC TTTGGGGAC	2160
TACATCCTCT CTAGTTCTT GCTTCGTGAG AAAGAATACT TTAAGTTGTC CGTAGGTCTC	2220
CAAACCTTCG TTAACAATGC GAAAACTTG AAGATTGCCT ACTTCTCAGC AGGTGCTATC	2280
CTCATCGCCC TTCCAATCTG TATTCTCTTC TTCTTCTAC AAAAGAACTT TGTTTCAGGA	2340
CTTACAAGTG GTGGCGACAA GGGATAATTT ATCCCCGCA CCCTTTTCA TTTTATACTC	2400
TTGGAATAAT TCTTCAAACC ACGTCAGCTT TATCTCCAAC CTCAAAGTTG TGCTTTGAGC	2460
AACCTGTGGC TAGTTTGCAC TTTGATTTC ATTGATTATT AGCAATTGTC ACTGTAAATA	2520
ATATCCTTGT AGCAAGCAAT TTTTCTCCTA GACTTGAAAT AAAGCGCATT TCTCTATATA	2580
ATAATACTCA TATAGAAAAC ACCTTTTAGA AAGATACCTA TGCTTCCATA TCCATTTTCC	2640
TACCTTTCAA GTATTGGGG GGTTCGTAAG CCGCTGTCCA AACGTTTCGA GCTCAACTGG	2700
TTTCAACTTC TCTTACCAG TATCTTCCTT ATCAGCTTGT CTATGGTACC CATTGCTATC	2760
CAAAACAGCT CCCAGGAGAC CTATCCGCTA GAAACTTTTA TCGATAATGT CTATGAACCT	2820
CTGACAGATA AGGTGTGCTA GGATCTCTCT GAACATGCTA CAATTGTGCA TGGCACATTA	2880
ACTTATACTG GAACAGCTAG TCAAGCCCCT TCTGTTGTGA TTGGTCCAAG TCAAATCAAG	2940
GAATTACCTA AGGACTTGCA ACTGCATTTG GATACAAATG AGCTAGTCAT CAGCAAGGAA	3000
AGCAAGGAAC TGACCCGCAT CTCTTACCGA GCCATTCAGA CTGAGAGTTT CAAAAGCAAA	3060
GACAGCTTGA CCCAAGCAAT TTCTAAAGAC TGGTACCAAC AAAATCGTGT CTATATCAGC	3120
CTCTTCTTAG TTCTCGGTGC GAGCTTCCTC TTTGGTTTGA ATTTCTTTAT CGTCTCTCTT	3180
GGAGCTAGCT TTCTCCTTTA TATCACCAAA AGATCACGCC TCTTTTCATT TAATACCTTT	3240

342

AAAGAGTGCT ACCATTTTAT CTGAACTGT TTAGGATTGC CGACTCTAT TACACTTATT	3300
TTGGGATTAT TTGGCCAAAA TATGACAACC CTGATTACTG TACAAAATAT TCTTTTGT	3360
CTGTATCTGG TCACTATCTT TTATAAAACA CATTTCCTGT ATCCAAATTA CCATAAATAG	3420
GAGATTTTAA TGCCCGTTAC GATTAAAGAC GTGGCCAAGG CTGCTGGTGT TTCGCCTTCA	3480
ACCGTAACCC GTGTTATTCA AAATAAATCA ACCATTAGCG ACGAAACAAA AAAACGTGTT	3540
CGCAAAGCTA TGAAGGAACT CAACTACCAC CCAAACCTCA ACGCTCGTAG CTTGGTAAGC	3600
AGCTATACTC AGGTTATCGG ATTAGTTCTT COTGATGACT CAGACGCCTT CTACCAGAAT	3660
CCTTTCTTTC CATCGGTCTT ACGTGGCATC TCTCAAGTCG CATCTGAAAA CCACTATGCC	3720
ATTGAGATAG CAACAGGGAA AGATGAGAAG GAGCGTCTCA ACGCTATTTT ACAAATGGTC	3780
TACGGCAAGC GTGTAGATGG GCTAATTTTT CTCTATGCCC AAGAAGAAGA CCCTCTCGTA	3840
AAACTCGTCG CAGAAGAACA GTTCCCTTTC CTTATCTTAG GTAAATCTCT ATCTCCTTTC	3900
ATCCCACTTG TCGACAACGA CAATGTTCAA GCTGGTTTTG ATGCGACTGA ATATTTTCATC	3960
AAAAAAGGCT GCAAACGCAT TGCCTTTATC GGAGGAAGTA AAAAGCTCTT CGTGACCAAA	4020
GACCGTTTAA CAGGCTATGA ACAGGCGCTT AAACATTACA AACTTACCAC TGACAACAAT	4080
CGCATCTACT TTGCCGACGA GTTCTCGGAA GAAAAGGGCT ATAAATTTAG CAAGCGATTA	4140
TTCAAGCAGC ATCCACAAAT TGATGCTATC ATCACAACCG ATAGCCTCCT AGCTGAAGGT	4200
GTTTGTAAC TATTGCCAA ACACCAGCTG GATGTCCCTG TTCTCAGCTT TGA CTGCTT	4260
AATCCCAAGC TCAACTTGGC AGCCTATGTC GATATCAATA GTTTAGAGCT TGGTCGTGTT	4320
TCCCTTGAAA CTATTCTCCA GATTATTAAT GATAATAAAA ACAATAAACA AATTTGTAC	4380
CGTCAATTGA TCGCCACAA AATTATCGAA AAATAAGAGA CTGGGCAAAA AGTCGTTAAA	4440
AGCAAAAACG CATACTATCA GGTATTGAAA AACTTGATA CTATGCGTTT TATTGTGGGA	4500
AGATTTACTT CCTTTTCTAC TGAAATTGAG TCTTTTCCCA AGATCTTTT ATACTCAATG	4560
AAAATCAAAG TGCAAAC TAGAAGCTAGC GCAGGTGCT CAAAACACTG TTTTGAGGTT	4620
G TAGATGAAA CTGACGAAGT CAGTAACCAT ACCTACGGCA AGGTGAAGCT GACGTGGTTT	4680
GAAGAGATTT TCGAAGAGTA TTAATCACTA ATTATCTATC TCAACAAATC TTCCTAGAAT	4740
ATGAACATTT TCCGAGACAG AGACAAAGGA GCTTGGATCC ACTTGTGTCA TAATCTGTTT	4800
AAATTCATTA AACTCTGCAC GTGTAATGAC AGTGATTAAA ACTGCCTTTC TCTCGTGATT	4860
ATAGGTTTCT TCTGCATCGT GGATCATGGT TGCTCCGCGG TGCAATTTTT TATGGATTTT	4920
TTCAATTACC TTCTCTGGAT GATTTGTAC AATCATGGCC TGCATACGCT TTTGCTTAGT	4980
AAAGACTGCG TCTGTACAC GGCTAGAGAC AAAGATGGTA ATCATAGAAT AAAGACCGTA	5040

343

TTTCCAACCA AAGGTCAAAC CTGCTATCAG CATGATAGTT CCATTTACCA AGAAAGAAAT	5100
ACTACCGACA TTCTTACCCG TTTTCTTACG AATAGTCAGG CTGACGATA7 CCGTCCCACC	5160
ACTGGAGATA TTGTTTCGAA GAGCAAAACC AATCCCCAAA CCCATAACAA CCCCCC AAA	5220
AAGGGAATTG ATAATGGGAT CCTCTGTCAA GGTGCCCACA GGGACAAACT GGATAAAGAA	5280
GGAACTCATA GATACCGTGA TAAAGGTAAA GACGGTGAAC TTATGGCCAA TCTGATACCA	5340
AGCTAAGACC ATCAAAGGGA AGTTAATGGC GTAGAAGCTT AGCGAAATCG GAATATGAAA	5400
ACCAAACCAG TGATTACTCA AGGCAGAGAT AATCTGTGCC AGACCTGTTG CACCACTCGA	5460
ATACACATGC CCTGGTTGGA AAAAGAAAT7 AACTGCTACT GCTGATAAAA AACCATAGAC	5520
CAGAGAGGCC GAAATCTTCT CATCATACTT TTCTCGAGAG ATACTTTGTA AGACACGTAA	5580
AATTTTTATC TGATAAGCAA AGCGGCGCAG ATAATAGCGC CACCGCTTAA TTCGTTTTGT	5640
TTGTTTCATC TTCTTCTACT TGTAAGCTGA GTTCTCTAG TTGTTTGAGA GCGACTGTTG	5700
ATGGAGCTTG TGTCAATGGG TCAGTTGCCT TGTGTTCCTT AGGAAAGCCA ATGACTTCAC	5760
GGATATTTTT TTCTCCAGCA AGCAACATGA CAAAACGGTC AAGCCCGATA GCCAAACCAC	5820
CGTGTGGTGG GAAACCATAG TCCATGGCTT CAAGAAGGAA ACCAAACTGG TCATTGGCTT	5880
CTTCAGTTGA GAAACCAAGA GCCTTGAACA TGCGTTCTTG AAGGTCTTTT TGGTTGATAC	5940
GAAGGCTACC ACCACCAAGC TCATAACCGT TCAAGACGAT ATCGTAAGCA ATGGCACGAA	6000
CCTTAGCCAA ATCACCTTCT AATTCATGAG CAGTCTCTTC CTGTGGAAGT GTGAAAGGAT	6060
GGTGGGCGCT CATGTAGCGG CCTTCTTCTT CAGACCATTC AAACATCGGC CAGTCAACCA	6120
CCCAAAGGAA GTTGAACCTA TCATTATCAA TCAAGCCAAG CTCTTTAGCA ATACGTCCAC	6180
GAAGGGCACC CAGTGTGCA TTAGCCACTT CAAGCGTATC CGCCACAAAG AGAACCAAGT	6240
CCTTATCTTC AAGAACAAGC GCTGTTGTCA ATTCTTCTTG GATACCAGTC AAGAACTTGG	6300
CAACTGGTCC GTTTAATCTT CCATCAACCA CCTTGACCCA AGCAAGACCT TTGGCACCAT	6360
ACTGTTTGGC TACTTCCGTC ATCTTGTCGA TGTCTTTACG TGAATAGTTG TCCGCAGCTC	6420
CTGTGACCAC AATCGCTTTT ACAGCAGGTG CTCTGAAAA GACTTTAAAG TCTACACCTC	6480
GGACCACTTC TGTCAAGTCC TGAAGCAACA TGTCAAAACG AGTATCTGCC TTGTCAGAAC	6540
CGTAAAGAGC CATAGCATCA TCGTATTTCA TACGAGGGAA TGGTAGCGTT ACTTCGATGC	6600
CTTTTGTTTC CTTTCATCAG CGCGCGATCA AGCTTTCTGT AATATCTTGG ATTTCTTGCT	6660
CAGTAAGGAA GGACGTTTCC AAGTCGACCT GAGTAAATTC AGGCTGGCGG TCTCCACGCA	6720
AGTCCTCGTC ACGGAAACAT TTAACGATTT GGTAGTAACG GTCAAAACCA GCATTATCA	6780

344

AGAGCTGTTT CGTGATTTGT GGACTTTGAG GAAGAGCGTA AAAATGCCCC TTATTAACAC	6840
GAGACGGCAC TAAATAATCA CGCGCCCCTT CAGGCGTTGA CTTAGAAAGG AATGGTGTCT	6900
CCACGTCGAT AAATCCAAC TCATCCAAGT AGTTGCGGAT AGAGTGGGTC ACCTTGGCAC	6960
GAAGTTTAAG ATTTCCAAC ATTTCTGGAC GACGAAGGTC AAGGTAACGG TAACGCAAAC	7020
GTGTATCGTC ATTTGCCTCA ATGCCATCCT TAATCTCAA TGGTGTGTG TTAGCTGTGT	7080
TAAGCACAAT AAGAGCTGTC ACGTTTAACT CAACCGCACC AGTTGGCAAC TTATCATTGG	7140
CTTGTCACGC GCAGCGACCT GACCAGTCAC CTCAATAACA AATTCGTAC GAAGGCTTTC	7200
AGCTGTGACC ATAACCTCTG CAGATACTTT TTCAGGGTTG ATAACCAACT GCATGATTCC	7260
TTACCGGTCA CGAAGATCGA TAAAGATCAA ACCACCAAGG TCACGACGAC GGCCAACCCA	7320
TCCTTTCAAG GTTATTTCTT GTCCGATGTG TTCTCACGA ACACGACCAG CATACTACT	7380
ACGTTTCATT ATTTCTCTCC TCTTTTATTC TGTACTATT TTACCATAAA AGCGCAGCTC	7440
TTCATGAAAA TCATCAGAAA AGTTTGCCAG TCTTTAAAAG TCAGGTGAAA GCCCTAAAAA	7500
TTAGCGCTAA TACTCTTCGA AAATCTCTTC AAACCACGTC AGCGTCGCCT TACCGTATGT	7560
ATGGTTACTG ACTTCGTCAG TTTTATCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC	7620
GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA CCTGCGGCTA GCTTCTTACT	7680
TTGCTCTTTG ATTTTCATTG AGTATAATAC AAAAATCCGA TGAACCTCAC CGGACTCTTT	7740
TATTTTGAAT TTTTGCCTGC TTTACGCTTT TCAGCGATTT CGGCTGCCTT TCGAGGCAAG	7800
ACAAATTTCCG TTATGTAAGC CGTCCCAAAA CGCAGTACAC CTGCAATAGG AGCAAAGACA	7860
ACTGCTAGAT AGTTATAGAA GAAATCGCCT TTGAAGGCA? AAGCTAGCGC TCCAATGATG	7920
AAAAATAGAA CGACTGCCTG AATCACTGCT AATAAAATTA CTCGTTTCAT GTGACCTCCT	7980
GACTCTAATA TAGCATGAGA ATCATCAAAA AGCCGACTAA ATTATTCAAA GCGTGAAGAG	8040
AAATACTGTA GACCAGACCT TTTCTGCTAA TGTAAGCCAA ACCCAAATA AAACCAAGGC	8100
TAAAATAGAC AAAAAATTGT TGCACATCAC CTGGAAAAATG AATCAAGGCA AATAGAAGAC	8160
TAGATACCAG AAGAAAAATC AGGGTTTCGTT TACTATTGTC CTGCTTAGGA AAGAGATAGC	8220
GTGCTAACAT CCCTCTAAAA ACAATCTCTT CCGTCAAAGG AGCAAAAATA ACCACAGCAA	8280
AGAATGAGAA AAGTGGTTGA GACAAGGTCA AGTCTGTCCG TATTTGCTGA TTTACTGAAG	8340
GATCATCTGG CAAGAAGAAT TGAACGACCA GAGATAAGAA CCAAACCAAG ACAGGAAGCC	8400
AAATAAATCG ATTAAGACCG CTCTTCTCAA TATGAACAGG AGCCTTCTGA TACCATTGTG	8460
AAATGCCGTA CACATATAC? CCAGCCAAGG CCACATAGAG TAGAGTAACA GCATAGGGTG	8520
AAGCGCCTAA AGCAAGCGAC GCAGTCGCGA GCCCCTGAAT AAAGCCATAG ATAAATAAAA	8580

345

AGGATAGAAG GGCTAGAAGA ATCCAGCCAA GGTTTTAAAG TAATTTTCATA GATAACTCCT	8640
TTATTTGAAA TAACGTTTTA CCATAGGTAA CTGCATCACA TTGATATAAA CATGGATGGC	8700
TCCTACAAGC AAGAAAGCTA GTAACCTGAAT CTCTCCTGTC AAGAAAGAAA TGATAATAAG	8760
AAAAATATAT AAGGCTGGTA AGACATATTG GTGTAATTGG AATAAAATTC GAAAACCTCTG	8820
TTCCAAATTA GCCTGACGCT CCCCTTCATC ATAAGAATTT ATATAGTTCA AGACATCCTT	8880
TGGTGTAGCG AAAAAATCCA AATCAAACCTG ACGAACAATC GCAATGGTTT TAAAAAGAGA	8940
TTTTTGAGCG ACTAAGAATA CCACAAAGAG TAAGAAAGAA AGGAAAAATG TTTGAGGGTT	9000
TGTATGCAAT ATAATCACCT CACTTAATGA AATAAAAATA GCCAATGGAA TCGCTACACC	9060
TGTAATATTA AAAGCAATGG TTCCAAACCTC AAGATTCCGA TACATTTGCA CATAATAGGT	9120
TTCATTGAGA TCGTCATCCA TTCTCCTCTG ATACAAAGAA TGAAATTTTC TGCTTTCTCT	9180
TAAGAAATTG AAAGTCAAAA ACATACTAAT GAAACCTATC AGTAAACAAA TAGCTGATAT	9240
CCATGGCATC AAGGCTTTTA CATCTAAAAT AATTTCGTGG GATTGACAC GTGCCTTAAA	9300
CATCCCTACA AACATGCCCC AGAACCCCC AAGACAATAG ACATCAAAAA TAACAATCTA	9360
CGTTTCTTTT TCATATTCAT TCTCCTTTT CACTTGCTAG ATTTTGGAT TTCTTTTCAA	9420
TCCATTCAAT TACTGGGATG AGAGCAAAGT AGACCCAAAC AAATTGGTCG CTTTGATAGG	9480
GATTAAACCA GCTTAGGTCC ATCCCAATCA GTAGAAATAC GCTGACTAAT AAAGCTATGA	9540
CCACTACATA ATAAATCACT TTATACTTGT TCATCACTCG TCCTCCTCCA AACGAAATAC	9600
CGATTGCACT GTTTCGTTGA AAATTTGAGA TATTTTCAGG GCAATGATAA TGGATGGGGT	9660
GTACTCATCC CGTTCTAGTA GGCTAATGGT CTGTCTGGAA ACCCCTGCCA GTTTGGCTAG	9720
GTCCGTTTGA TTGAGACCAT CCGGAGCTCG AAGCTCTTTT AGACGATTTT TTAGTTGCAT	9780
GTTACACACC TACTCTCCGT CAAATTCAAC GGTTTGGATA TCCTCAATAC GTTGCAACTT	9840
GAATTTTCT TTTCCCGTAT TATCTACAG TCGTAGCTT ACCCATTCCT CATCAACATC	9900
CACAACTTCC CAGTTATCTG GCCCAATATA CACTCCCGTT ATAATTGGTT CCTTTCCAAT	9960
CATTTCTTGT AATAATCTCG ACATTTCTGC GTTTCCTTTC TCTTTTCGCT CAAGTCTTTT	10020
GATTTTATTC TCTAGTTTCT TGATTTTTTT AGAATTATTA GAATAAAGA AAATCATAAA	10080
TAGTATAAAT CCTAGTACCC ACATTATAAC TCCTTTCTGC TTCCTATTTT TTAACCTGAA	10140
TTCAATGTAA CATATCTTTT TCTTTTGTGAC AAGTATAGTT GTCAAAAAA TTATGATTTT	10200
TGTCATTTTG CAAAAGAAAA AGGTCAGGAG TAGGTTCCCTG ACCACTTTAT CTATCATTAA	10260
TACTCTTCTA AAATCTCTTC AAACCACGTC AGCTTCACCT TGCCGTAGGT ATGTTTACTG	10320



346

ACTTCGTCAG TTTCATCTAC AACCTCAAAA CCATGTTTTG AGCTGACTTC GTCAGTTCTA	10380
TCCACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTCTATCCAC AACCTCAAAA	10440
CCATGTTTTG AGCTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTTGAGCAA	10500
CCTGCGGCTA GCTTCCTAGT TTGCTCTTTG ATTTTATTG AGTATAAAAT CCTAGTTTTT	10560
CAAAGATTTC TGAGAAGTTT TGGCTGATTG TCTCAAGTGA CACTTGCACT TCTTCTCGGG	10620
TTTGGTTGTT CTTGACCGTC ACTTGTCCGC TTTGACTTC GCTCTCTCCT AGGGTGATGA	10680
GGGTCTTAGC CGCAAAGACA TCGGCTGACT TGAAGTGAAG TTTTAGTTTA CGGTTGAGGT	10740
AATCAGGCTC TGCTTTGAAA CCTTGTGGC GAAGAGCCTG TACCAATTCC AAGGCCTTGA	10800
TATTTGCCCC TTCGCCCCAG ACTGCGATAT AGACATCTAG GCGGTTTTCG ATAGGGAGGG	10860
TCACACCTTG CTTTCAAGG ATGAGAAGCA GCGCTCTAC ACCAAGTCCA AAACCAATC	10920
CAGCAGTTTC AGGGCCTCCA AAGTAAGCAA CCAAACCATC GTAGCGACCA CCCGCACAGA	10980
CGGTCAAGTC ATTGCCCTCA ATCTCTGTGA TAAACTCGAA AATGGTGTGG TTGTAGTAGT	11040
CCAGACCACG CACCATATTG GTATCGATGA TGTAATCTAC TCCAAGATT TCCAACATCT	11100
GACGCACAGC ATCAAAATGA GCTTGGCTTT CTTCATCAAG AAAGTCCAAG ATAGACGGCG	11160
CATTCTCTAC TGCCACCTTG TCTTCTTTTT CCTTAGAGTC CAAGACACGA AGAGGATTTT	11220
CCTCCAAGCG ACGTTGGCTA TCCTTAGACA AGGTCTCCTT GAGCGGTGTC AAATAGTCAA	11280
TCAAGGCTTG GCGGTAGGCT GCACGGCTCT CAGGATTTCC AAGAGTGTG AGGTGCAATT	11340
TGACACCTTG AATACCGATT TCCTTCAAAA AATGGGCTGC CATAGCGATT GTTCCACAT	11400
CGGTAGCTGG ATTGCTAGAG CCAAACACT CAACACCAAT CTGGTGGAAT TGGCGCAAGC	11460
GCCCTGCCCTG TGGACGCTCA TAACGGAACA TAGGTCCCAT GTAGTAGAAC TTGCTTGGCT	11520
TTTGCACTTC TGGGGCGAAA AGTTTATTTT CCACATAGGA ACGGACAACG GGTGCAGTTC	11580
CTTCTGGACG GAGGGTAATA TGACGGTCAC CCTGTGCTA AAAATCGTAC ATTTCTTGG	11640
TTACGATATC CGTTGTATCT CCGACAGAGC GACTGATAAC CTCGTAATGC TCAAAAATAG	11700
GCGTGCGCAC TTCTGCATAG TTGTAGCGTT TGAAAATCTC ACGGGCAAAG CCTCAACGT	11760
ACTGCCACTT AGCAGACTCA GCAGGTAAAA TATCCTGCGT TCCTTTTGGT TTTTGTAAAT	11820
TCATAGGGAA TCCTCTTTAA ACTTAATAGT CTTATTTTAC CATAAATAGA GGGATTAAAA	11880
CAGTAAGAAA AAAATTAGGA TTTAGATATC ATTTTGTAGA TTAAGAATTG TCAAAAAAAT	11940
AGCTAGCAAG GAAAGACCAA CAAATAGCAT CCAAGTCAAC TGTATATTCC ATACGGCTAC	12000
TAGTGAAAAA CAAGCTGTTC CCACAGGTAT GGATAAGGTA AACAATAGAC CTAATAAATT	12060
ACTAGTACGA GCTAGAACCT CTGGAGCTAG ATTTTTCATG AGCATGGCAC TAATCTTGG	12120

347

TTGAACTTTA CCAGACACAT ACAGAGTAAA GAAGAGAAAT AGCAAACCAA GCACGACTTG	12180
ATTGAATAAA TTAGCCAAAC CAACTAGACT AAGTCCTACG GTCTCCCACA TCATCAATCT	12240
AGGCAAGGAC TGCTTCCCAA AATAATCATT GCCCGTAAGG CTA CTGATGA TGACTGATAC	12300
TAAAACACAG AATTGATTGA TAAATAGTGC CTCTGTATAA <u>GA</u> AAAAATTCA AGAGAGAATG	12360
GCTCAAAAAG AAGATATTAT AAATTCCACC CAAAGCGCCA CCCAAGGAAT TAATAAGCAA	12420
GACAGCAAAG AGCATAAAAC CAAAGTTTTT CTGTCCACTT TTAAGAAAAA CGAGACGTAA	12480
ATTTCCGTAA ATTGTTAGGA ACTGGTCTTT GATAGAAAGC TTCTCATTTT TTAAGTTTTT	12540
ACCATCAGCA GATGACATTG ACAGGCTCAA TTTGCTTTTT CCTAAAAAGA GGATAGTGGC	12600
TGATACTAGG AAAAAGCAGG CATTGATTCC CGCAACGAGA GAAAAATTGT TGACCGATAG	12660
AGCTAAGAGC CAGACTCCGA AAGCTTGACC ACCAATAGCT GAAATATAGG TGATGAACTG	12720
TGAAAAAGAA TAAGCCTCCA TCAGATCATC TTCAGCTACT TTTTCCTTAA TAAGAGGCAT	12780
ACGCAGGCCA CCTGCAAAAT CACTGATGAT ATCACTAATG ACATTGATCA AACACAGGCT	12840
AGAAAAGGCA AAGAGACTAG CTTGCTGAAC AACTAGGGCT GCTAGAAAAA ATAGAACC GC	12900
CTGAAACAAA CCGCTATAGA CCATCCATTT GACCTTGTC CTCGTGTAAT CTGCCCCAAT	12960
CCCTGCAAAA ACTGTAAAGA GGGTCGGAAG AATCATGACA ATATTCGCCA TAGCAACAGC	13020
AAAAGATGCT TGTGACAAGG TCGATGCATA GACGATAAAG ACCAGGTTGA AAATCGAAAC	13080
ACCAAAGCA TTGAAGAAGC GTGG	13104

(2) INFORMATION FOR SEQ ID NO: 35:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19250 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCGGGCAAAT AGTTTTGAAC TTTTCATCAT TTTCTCCTTT AAAACTTTCT CTCCATTATA	60
GACTCTTTTC AGAAAGTTGT CAACAGAATT TTCAGAATTT TTGAAAATTA TTTTCAAAC	120
AACATCTTTG CAAAAAATAT GAATATCGTA AGCGCGTCAT AACAAGGTAT CTATCATTCA	180
TGGAGCTCCT CTTGTATACT ATTAGTAAAG TAAATATTGG AGGATATTTT AATGCCACAA	240
CCTATTGTTT CTGTAGAGAT TCCACAATCT CGTCGTTTTG ATTCTAAAAA GAGAAATGAT	300
ATTCTTCTTA AAATTCGTAT TGGCAAGCTT GAAGTAAGTT TTTTCAATC TCTCAATCTC	360

348

GAAATGATAG AACAGCTTTT GGATAAGGTG TTGCTCTATG ACAATTCATC TATCTAGCCT	420
AGGGCAGGTC TATCTCGTGT GTGGGAAAAC TGATATGAGA CAAGGAATCG ATTCACTGGC	480
TTATCTCGTT AAAACCCACT TTGAATTGGA TCCTTTCTCC GGTCAAATCT TTCTCTTTTG	540
TGGTGGACGT AAAGACCGCT TTAAAGTCCT TTAAGTGGAT GGTCAAGGAT TTTGGCTACT	600
ATATAACGC TTTGAGAACG GCAGACTGAC TTGGCCCAGT ACAGAAAAGG ATGTCAAAGC	660
TCTCGCACCT GAACAAGTAG ATTGGCTGAT GAAAGGCTTT TCTATCACTC CAAAAATATA	720
GTAGATTGAA ACTAGAATAG TACACCTCTG CTTCTAAAAC ATTGTTAGAA ATCGATTTTA	780
CTGTCCTGAT CGATTTGTCC TGTATTATT TCAATTTACT ATAAATCCAT CAGAAAGTCG	840
TGATTTCTAT TGAAATGAGG ACTTTCTTTT TATACTCATC TGCTTTCAA AAGCACTCTA	900
GTCCATCTCC GATTAAACGAT GGACTTTATC ACCTCCTTCT CCAGTCCTTG TATAACATCT	960
TGAAGTTGAT TCATGACATC TTCCAAAGTT CGAAAGGCTT TATTCTTAAA TCCACGTTTA	1020
CGAATCTCTT TCCACACTTG TTCAATGGGG TTCATCTCTG GTGTGTATGG AGGAATAAAT	1080
GCAAAGCCAA TATTAGTCGG AATCTTTAAG GTACTTGATT TATGCCATAT AGCATGTGCC	1140
ATAACGAGTA AAAGATAATC ATCTGGATAA GCTTGTGAAA GCTCCTATTC CTAAAGCCCC	1200
TTTATAACCT CTTGCGAGAG AGACTATTGA CTCAGCCCTT ACTTCATGCG GATGAAACCT	1260
CCTATCGGGT TCTAGAGAGT GATAGCCATC TGACCTACTA TTGGACTTTT TTGTCAGGTA	1320
AAGCAGAGAA ACAAGGGATT ACGCTTTACC ACCATGATCA GTGTGGAAGT GGTTCAGTAG	1380
TACAAGAATT CCTAGGAGAT TATTCTGGCT ATGTTTCATTG TGATATGTTG CCGCAGTAAC	1440
TTAGGACTTT AGTCCTCTAG TTCTGCCTAT GCGATAGCAG TCCAAGGTTT AGGAGTAAGG	1500
CGACGCTAAG CTTGGTAAAC TGCGAACAGC TAGAAGCTTA TCGTCAACTG GAAGAAGCTG	1560
CACCTGTTGG ATGTTGGGCG CATGTGAGAA GGAAGTTTTT TGAAGTGCCC CCCAAGCAAG	1620
CAGATAAATC ATCCTTAGGA GCTAAAGGTT TAGCCTATTG TGATCAGTTA TTTTCCTTGG	1680
AAAGAGACTG GGAGGCTTTG CCAGCTGATG AACGGCTACA GAAACGTCAA GAACATCTCC	1740
AACCCCTACT GGAAGACTTC TTTGCTTGCT GCGGTCGTCA GTCAGTTTTA TCGGGTTCAA	1800
AACTAGGAAG GGCAATTGAA TACAGCCTCA AGTATGAAGA AACCTTTAAG ACCATTTTAA	1860
AAGACGGACA TCTGGTCCTT TCCAATAATC TAGCTGAACG CGCCATTAAA TCATTGGTTA	1920
TGGGACGGAG TAAAAGAGTC CAGTGGACTC TTTTAGCCTA AGCTCAGTTT AAAAAACGA	1980
GGGTGTTAT TTTTAAAAA GCGAGGGTGG TTATTTTCTC AAAGTTTTGA AGGAGCTAAA	2040
GCAAGAGCTA TTATTATGAG TTTGTTGGAA ACAGCTAAAC GTCATCAATT ATAGTGCCTT	2100
GAATCTATAA CAGTACGAT CGACTGCTAA AATATTTCTA TAAATCAATT TTCCTTTCTT	2160

AATCGATTTG TTCATATCTT ATTACAATCC ATTATAAATA GCGAGAAATA TCTATCCTAT	2220
CTTCTAGAAT GTCTTCCAAA CGAGGAAACT CTCGTAAACA AAGAGGTTTT AGAGGCCTAT	2280
TTACCGTGGA CTAAAGTTGT ACAAGAAAAG TGCAAATAAG AAATCTCCAG ATTAGGAACT	2340
ATATATGAGT TCTCTAGTCT GGAGATTTTT CAATAGACTT CGTTATTGGG CGGTTACTTT	2400
CGAAACTTTG AAAACTTCAA AAAACGGATT TTTATCGCTC TGAACATCAA AAAAGAAAGG	2460
ACGAAATTTG TCCTTTCTCA AGCTTAGCTT TTCTTCAACC CACTACAGTT GACAAAGAGC	2520
CCTTTATTCT ATCAAACATG AAGCGCAAAA ACAAGCCAAA AATCCGATAG AATGGCTATC	2580
CCTCGACTAT CAAGTAAGAC ATTTCCATCA AATACGTTCA ATTTTACTCT TGTCTACTA	2640
AGAATTAATC ATCTCGTTTT GATTTATTAA AAATATACAA TTCAGCTTTT CCTCCAAACT	2700
ATTTTATCCA CTATCCCTGT ATAGCTCTGT ATTATCTTAA CAACTTTAGT AGAGACATTT	2760
TCCTCAACAT AATCCGGAAC CGGTAATCCA AAATCCTCAT CTTGTGCCAA GCTAACAGCA	2820
GTTTCAACTG CTTGAAGAAG AGAATTTTCA TCAATGCCTG CCAAAATAAA TCCTGCCTTA	2880
TCTAAGGACT CAGGACGTTT TGTACTTGTA CGAATACATA CAGCGGGAAG AGGATAACCT	2940
TGACTAGTAA AGAACTACT TTCTTCCGGT AAAGTTCCCG AATCAGATAC TACAACAAAT	3000
GCATTCATCT GTAAACAATT ATAGTCATGG AATCCTAGTG GCTCATGCTG AATCACACGT	3060
TTATCTAGTT TAAAACCGCT CTCTTGTAGC CTTTTCTTTG ATCTAGGATG GCAAGAATAT	3120
AAGATTGGCA TATTATACTT TTCAGCTAAT TGATTAATTG CTGTAAAGAG AGAAATAAAA	3180
TTTTTATCTG TATCAATATT TTCTTCACGG TGAGCTGAAA GTAAGATATA ACCTCCTTTT	3240
TTCAATCCCA AACGTTTCATG GATATCTGAA GACTCAATAG CAGATAAATT TTTATGTAAC	3300
ACTTCTGCCA TAGGAGAACC AGTTACATAT GTGCGCTCTT TAGGTAAACC AACTCATGT	3360
AAATACTTAC GTGCATGTTT AGAGTATGCT AAGTTAACAT CTGAAATAAC ATCAACAATC	3420
CGACGATTAG TCTCTCCGGT TAGGCACTCA TCTTTACAGC GATTGCCAGC CTCCATATGA	3480
AAAATTGGAA TATGTAAACG CTTGGCAGCA ATAGCTGATA AACAAGAATT TGTATCCCCT	3540
AAAATCAATA AAGCATCTGG TTTAATTTGA TTCAATCAATT TGTATGAAGT ATTAATAATA	3600
TTCCCTACAG TAGCACCAAG ATCATCTCCA ACAGCATCCA TGTATACGTC CGGAGTGTCT	3660
AACCCTAAAT TATCAAAGAA AATACCATTT AAATTGTAAT CATAGTTTTG TCCAGTATGT	3720
GCCAAATAA CATCAAAATA CTTTCGACAT TTAGTGATAA CACTACTTAG ACGTATAATC	3780
TCTGGACGTG TTCCACAAT AATCAATAAC TTAAGTTTGC CATTATCTTT AAAGTGAATA	3840
TCACTATAAT CTGTCTTAAT TTTCAATTTAT TTCTCCACTT GTTCAAAAAA AGTATCTGGA	3900

350

TGTCTAGGAT CAAATGACTC ATTAGCCAC ATGACAGTAA TTAGATTTTC TGTATCAGAA	3960
AGATTAATAA TATTATGTGC ATAGCCCGGT ATCATATGTA TTGCTTCAAT CTTATCGCCC	4020
GACACTTCAA AGTTCAGAAT AGGATACTCT TGACCGTTTT CATCCAGCCC TATCCTACGC	4080
TCTTGTATTA AAGCACGACC AGAAACAACC ATGAAAAATT CCCACTTAGA ATGATGCCAA	4140
TGTTGCCCTT TGGTAATGCC AGGTTTAGAA ATATTAACAG AAAATTGACC CGTATTTTCT	4200
GTTTTTAATA ATTCCGTAAT ACTACCTCGT TCATCTATAT TCATTTTTCAG AGGAAACTTA	4260
AACTTATCTA CTGGTAAATA AGATAGGTAG GTAGAATACA ATTTCTTTTT AAACGATCCC	4320
TGAGGAATTT CAGGCATAAC TAACTATCA GGCTGTTTTT TAAATGTTTC TAATAGAGAG	4380
ACAATCTCTC CTAAGGTTGC ACGATGAGTC GTTGGTACGT AGCAGTAGTT TCCTGATGGG	4440
CTAGGTAAGA TTTGTAATCC ATCTAGATTA CAACGATGAG GATTTCCTTC CAATGCAGTT	4500
AGACACTCTT GTATCAAATC ATCAATATAC AGCAACTCCA ATTCTACACT TGGATCATTT	4560
ACTTGAATAG GTAAATCGTG AGCTAGATTA TAACAGAAAG TTGCTACAGC AGAATTGTAG	4620
TTAGGACGGC ACCACTTCCC ATAAAGATTC GGGAAACGGT AACTAAGAC AGGTGCTCCC	4680
GTTTTCTTTC CATATTCAAA GAAGAGTTCT TCCCCTGCTA GCTTAGATTG TCCATATATA	4740
GAGTTTGAAA ATCGGCCTTC TAACTAGCT TGAGTAGAAC TTGAGACTAG AACAGGACAA	4800
GTGTTTTCAT ACTTTTCTAA AATCTCCAAT AATCTACTTG AAAAACCGTA ATTTCCCTCC	4860
ATGAATTCAT CAGGATTCTG TGGACGATTG ACACCAGCTA AATGGAATAC GAAATCGGCC	4920
TTCTTACAAT ATTCATCTAA TAAAATCGGA TCTGTATCAC GATCATACTG AAAAATCTCT	4980
CCAATCTCTA AATTAGGACG AGTCCTATCT CGTCCATCTT TCAAAGCTTC CAGAGTACAG	5040
ATAAGATTTT TTCCTACAAA TCCTTTCGCT CCTGTGATTA AAATATTTTT AATCATGCCC	5100
CCTCCTTATT TTATATGCTC TTTTAATAGT TAACTCTCTC GACAATACAT GATACATTAT	5160
ATATCCTTGA TAATTTTAAT GTATCTTAAA AGATTTTACA TCTCTTCGTC TGCTACCATA	5220
TCACGAATTG CTGTCTGTAT TTCATCTAAT TCTAGCAACT TTCTTTTAAC TTGCTCTACA	5280
TCCATCAAAT CGGTATTATT ACTATTGAAT TCTGTCAACA AATTTCTATT CGTACTACCA	5340
TCTTTGAAAT ACTTATCATA GTTAAGATTA CGATTATCAC TAGGAACTCT ATAAAAATCA	5400
CCCAAATCAA TTGCATTTGC GCACTCTTCG TTAGTTAATA GTGTTTCATA CCTTTTTTCT	5460
CCGTGTCTAA TACCTATAAT CTTAATATCT TGTTCTGAGG CAAAAATTC TGATACAGCC	5520
TTAGCCAACA CTTCAATCGT ACATGCTGGT GCTTTCTGAA CTAGTATATC TCCAGATTTT	5580
CCTTCTTCAA ATGCAAATAA AACCAAGTCT ACTGCTTCTT CCAATGTCAT CACAAAACGT	5640
GTCATGCTAG GTTCAGTAAT TGTAAGAGCA TTTCCTTGCT TAATTTGCTC AATCCAAAGA	5700

351

GGAACGACAG ATCCACGGCT ACACAGAACA TTCCCATAGC GAGTCACACA TATCTTTGTA	5760
TGCTCAGGAT TTACCGTCCT GGACTTAGCA ACAGCAATCT TTTCCATCAT AGCCTTGGAT	5820
GTTCCCATAG CATTGACAGG ATAAGCCGCC TTATCTGTAG AAAGACAGAT AACTTGCTTT	5880
ACACCAGCTT CGATAGCCGC AGTGAGGACA TTCTCCGTTT CCAAAATGTT AGTTTTTACC	5940
GCTTCTACAG GGAAAAATTC ACAAGAAGGT ACTTGTTTAA GAGCAGCAGC GTGAAAAACA	6000
TAATCCACAC CATGCATAGC ATTTTTTACC GAAGCTAAGT CACGCACATC TCCAAGGTAA	6060
AAACGGATTT TCCCAGCCAC TTCTGGTACT TTTACCTGAA ACTCATGACG CATATCATCT	6120
TGTTTCTTTT CATCTCGCGA AAATATACGA ATCTCTGAGA CATCTGTTTC TAAAAACGC	6180
TTGAGAACCG CATTCCCAA TGAACCTGTC CCTCCTGTAA TTAGGAGAGT TTTTCTGTA	6240
AATTGTGACA TATATTACAC TTCTCCTTCT AGTATGTCTG CAATTTTCTT ACAAGCCGTT	6300
CCATCTCCAT ATGGATTGA AGCTTGACTC ATTGCTTGAT AAAGTGAATC ATTTTCTAAT	6360
AATCTTTTAA AATGCCTATA AATATTATT TCATCAGCAC CTACAAGTTT CAAAGTCCCT	6420
GCTTCAATTC CCTCTGGACG TTCAGTTGTA TCTCTCATAA CCAAAACAGG TTTTCTAAA	6480
CTTGAGCCT CTTCTGAAT ACCACCACTA TCTGTTAAAA TTAAATAACT TCTTGATAAA	6540
AAATTGTGAA AATCTAATAC TTCTAAAGGT TCGATCATCT TGATACGTTT ACAGCCACTT	6600
AGTTCTTCCT CAGCAATTTG GCGAACACGA GGATTCATAT GGATAGGATA AATAGCCTTG	6660
ACATCTGAAT ATTCTTCAAT AATCCTTCTA ATTGCTCTAA ACATATGTCT CATCGGTTCA	6720
CCAAGATTTT CACGACGATG AGCTGTAATT AGAATAAACC TGCTTTCTCC TATCCATTCT	6780
AACTCAGGAT GCGTATAGTC CTCTTGAATT GTAGTTTGTA AAGCATCAAT CGCCGTATTA	6840
CCTGTCACAA ATATGCTCTC TGGAGTTTTT GCTTCTCTTA AAAGATTATC TTTTGAAAGT	6900
TGTGTTGGTG TAAAATGATA CTGAGCCAAA ACCCCAACTG CTTGACGATT AAATCTTCA	6960
GGATATGGTG AATAGATATC GTAAGTGCGC AAACCAGCTT CAACATGACC AATTGGAATC	7020
TGTAAATAAA AGGCCGCCAG TGAAGTAGCG AAGGTCGTAC TTGTATCCCC ATGAACTAAC	7080
ACCAAATCAG GTTTTTCTGA CTCTAAAATA GCCTTCATTC CTTCCAAAAT GCCAATGGTC	7140
ACATCAAATA AAGTTTGTTC ATCTTTTATA ATAGACAAAT CAAAATCGGG AATAATCCCA	7200
AATGTGTCCA AGACCTGATC CAACATTGA CGGTGTTGGC CCGTAACGCA AACTAATGTT	7260
TCAATATTCT TACGTGTTCT TAACTCTTTG ACCAAAGGAC ACATCTTGAT GGCTTCTGGA	7320
CGAGTTCCAA ATACTACAAC TACTTTTTTC ATATATTAC TTAATCCTAA CAAATAATGA	7380
ACGGTTCTTA AAATAAATTA GATAACGGCT AATCCATAAC ACCACCTCAG ACATACTTGA	7440

352

ACAAATAGCT AATGTTACTA AACTAAAATT ATCAGACAAG ATAAATATTC CTAATCCCAA	7500
AGTTTGGACA ATCGAAGCTA ATATAGTTGT CATTGTAGTT TCTTTCACCT TATCAATAGC	7560
TCCTAAGACA GGCCATCCGT AAATCATAGA ATAAAACTA GCAACAAAAG CGGGTAATAA	7620
GTACTTAAGA AAATCTGCTG AAACGGTATA TTTTTCACCA CCAATTATAG AAAGAATTTG	7680
ATTTGAAAAG AATAAACTA TCAAACTCC AAAGATAATA GGAATAAACA TAATCCGATT	7740
AATACTCTTA ACCGATTGTA TATCTTTAGT ACGTATCATA TGCGGATATA AACTATTCCG	7800
TATAGGATTA TACAATGATT TTGCTGCTGA AAGCAGTTGC ATTGCTATCC CCCAAAAGGC	7860
TATCTCTTGA CTTTGTAAT AAAAACCCGA AATGACTGTC GTAAAGACGC CAAAAATAGT	7920
AGTTGCAAAA TTGGATAAAA AATAAATAGA GGATTCTTTT AAATCTTTAA CCCAACAGA	7980
CAGATAAGAA AATGATAATT TAATCCATA ATAATGAAGG AATCTATAAG AACTACTGCG	8040
AGCAACTAAA TTCCCAATTC CTTCCAATAT AGGAATCCAT AAAATAGAAG AATCATCTTT	8100
TACTACAATA AATGTCAAAA TTGTAATGAT AGTTTTAGAA ATAATATAAG GAATTGCAAC	8160
TGCATGCATC TTTTCAATTC CACGAAATAA AAAGTCAAAG ATAAAAATAT TGGTCACTGT	8220
AGCTAACAAA TAAAAAACTG AAAAAAGAAT ATTCTCTCTC ATTATTGGGA TTTGCCACAT	8280
CAATATGGTG TAAATTAGAA TCGAAATGAT AGATAAAAAT ATTTTTTCAA CTAGAGTATC	8340
TCCAACTATC CTTCCAATCT TTGAGGGAGT AGTACAAGCA TTTACAATAT TTTTGTAGC	8400
TGATATCATG AAACCAAAAT CAATCACCAG TTGAACATAA GCTATTAAACG CTTTAACATA	8460
AATAACCATT CCATACGCGT CTAGCGAAAG CACCCTTGTC AAATACGGGA GTGTAAATAA	8520
AGGAAATAGT AATTTAACAA TATTCAGAAT ATAGAGAGAA CTTGTATTTT TTATAAATGA	8580
AATTCTATCA ACTTTCACGA ACTAGTCTTT CCAAAAAAAG ATCTAAATAG TCCAACTAC	8640
TTCTCGCTTT CAACACCAAT TCTGAAGGTA TTGTTATCGG TTTTAGATGA AAAGTTTCAA	8700
GTTTCTTTAC AATACTATTA ACACCTGAAT CAAATAAAGA TTCACAACGT TGTAACCTC	8760
CAATTGCTCC ATAATAACGT GCTGTTTTTT CTGGATGGCA TGCAATGGCA ATCACAGAT	8820
TATTAAAACA TGTTGCCACT ACCCCAACAT GTAATTTACA AGTTAAACC ACATCTACCA	8880
TTTTCAACAA TGATGTCATT TCTGCAGGAG AATGATACTT GAATTGAAAA CAATCCTCAG	8940
TTCTAACTAA TTTTCTAAAT TCCTGATAAT AAGCATCTTC ATAAGGTAGA ATGGAATCCG	9000
AAGTTACTAC AACATAATAG TTAGGATTGT TTTCTAGAAA AAGACTAATT GATTCCGCAA	9060
ATTTTCAAG AGCTTTTTTG GAATGATTAT AGTGAACAAG AATTATCTTC TTATCTTTAG	9120
CTTCTCTTTT CAATTGACAC AGCTGCTCTG TTTTCTTC TCTTAATTTA CTTGAAATAA	9180
TTAAATCAAA GGTTCATGC ACTGGAGCCG AAGGCGACAA ATGCTTCAAA GAATCAAATG	9240

353

ATTCTCGATC	ACGAACTGTA	ATAAATTGAG	CATGATTAAAT	AATTCTCTTT	ATACCATAAT	9300
TCATCAAAGA	ATCGTTATTA	GGCCCTGCAC	CAATACCTAA	TACTCCTATA	GGCTTTTTAA	9360
AATATGAAGC	CCAAATCCCC	AAAGGTAAAA	ATCGTTTAAA	TTGGATTAAA	TTATCACGAA	9420
AACGTGCATT	ATGCCCTTCC	CCAAAATATC	CTCCCGGGAT	ATACAAAATA	GCATCTGCTT	9480
GTTTTTTAGT	AAAACTTTGT	TTTTGGCGAT	ATTCTTTCAA	GTACATTTGA	AAGAAATCTG	9540
ATGGATTATA	AAAAGAAACT	TCATATCCTT	TAGATTCTAA	TAAATCATAG	ACAATCTCAC	9600
CGTAAAGATA	ATCACCGTAA	TTACTTGAAC	CATAATCCGT	TGCACCATGT	AACATAATTT	9660
TTTTCACCAC	TATTTTTTCA	ACCTCCTAAA	AATAAATATC	ATAATCAAAC	TATACATAAT	9720
AGGACGATAA	ACATCTATTG	AACTACTTCT	CACTAAAAGC	AATAGTTGAG	AAATTACCGA	9780
AAAATAAATA	ACTTTTGAGA	TTTACTTGT	TTGAAAAGCT	CTGAAATTTA	ATCGCCATCC	9840
ACTAAATATT	CCCCAAACAA	AACCTCAAAA	AACACCACCA	TAGTAACCAA	AGTTCCAAAA	9900
TAATTCTTCC	ACAAAAGAAG	AGCCTACAGG	TAACCCCAAA	AATTTATTAA	TAACAACCGT	9960
CGCTGATGCT	TTATCAAAAA	AATCACCAAC	TAACCATCCA	ATAGGAAAAA	TTGATAGGAT	10020
AGTCCGTAGA	AATGTCATCC	CATATTCATA	TGGAATGCTA	CTAGGCACAA	CAGTTACAGC	10080
AGAAGCTACT	GTTAGGCTGG	TCAGTCCCGA	CTCTGAAAAT	ACTTCCCCTA	GTATATTCTT	10140
TACAAAATCT	AATGAAGAAA	AGGAATCAAA	TAAGTATATA	CCTATAGTAT	TCAAGTCGAA	10200
ACGGTGCCCC	CTAATAACAA	CTAATACATT	TAATAGAAAT	ACAGTTACTA	TTAAAAATAC	10260
AAGTACTCTT	TTCTTCGAAA	AAGTAATCCC	TAAAGATTGT	GTGTATACTA	AAACCAACGC	10320
CAAGATTGAA	AACACCTGGA	TTTTACGACT	TCCTGTTAGG	ATCATTATCA	AAATTAGGTA	10380
AAACAACATT	ACCCAAAAAA	TAGTACGCTT	TATAACTCGG	GACAGCTTAT	CTGAATAAAA	10440
CAAGGAGAAC	ACACCAGGAA	GCATAAGTAC	TCCTAAATCA	TCTATTATTC	CTGAACTAGC	10500
TGCCCTTGAA	TATGCTGAAT	AGCTATTGCG	CGCTCTAACT	GCTAGTACTG	TTTTAGAAATC	10560
AGTTATTACC	CTAGAAATAA	AGCCCACTCC	TGTTAAAAATC	CTACCCGCAT	TGTACAAAAT	10620
TTTCTCTTCA	TTTTCCTGAT	AATTTGTAC	TTCTGAATGA	TAATGTACCT	TTCCATCACT	10680
ATAAAAAAAT	AAATAGCCTA	CAGAATAACA	AAACAAAATC	CAAATTATAA	AAATATATGA	10740
ATGAAATAAT	TCTTCATTAT	TATAGAAGTT	ACTAGGGCTC	CACAGCAGAG	TTGTTTGAAA	10800
CCCCATATAC	TCATTGAAAA	TTAATCCAAA	CATAAAAAAA	TAAGATAAAA	TCAGATACCA	10860
TACAGAAAAA	TCATATATAC	TAACTTTTTG	TAAAATAAAA	CCAGTAATTT	GAAAAATAAT	10920
TAGAAAGCAA	ACCCATATAA	ATATAGACGG	AACATAATTA	GATATAAGAA	AACCATTATT	10980



354

CCAATTATCG AGAGTCCAGA ACAAGTAACA GAAAGCAAAT <del>ATAAA</del> ACTTA ATGTCACTAG	11040
TGTCACTCTA CAAATATACT TTGTCTGCAT CTATATCTCC TTTATTACAC ACATTTCTTG	11100
ATAACGATTC AATAATTAC TAGCTTGATA ACAAATATCA TAGAGTCCAT CTGTCATACT	11160
GTTATTTATT TCAAAACGAT TGCATTCCCTC AGATGTTAAA GACAGTACTT TATCTTTCCA	11220
TAGCAACACA GACTCTTCGT TGATAGGTAA GTAACATAATG TTTTGGTCA CATCTACTTC	11280
TTGCGTCACT GTATCTGACG ATAAAATTTG TAATCCCGAT GCCTGAGCCT CTACTAGAGA	11340
AACAGGCAAC CCCTCATATT TAGACGGAAG CAAAAAACA TCCATCGCAG ATAATAAATC	11400
AGAAATATCA GTCCTTCTCC CTAAAATAG CACATATGGG GTCAGATTTA GTTCTAAAGC	11460
TTCTGTGTTT AATTTCTGCT CATCCTCACC ATTACCAACT AGGAGTAAAA TAACATTTGG	11520
TTTGATTAAA ATGAGTTCTT TTAACCGTT AAATAAATAA CTTTGGTTTT TTTGATCTGA	11580
TAGGCGAGCT ATATTTCTTA ATACGAACTT ATTTGACACA TCTAATCTC TACGACATTT	11640
TTCTCTAACA TCTGACAAAA ATTGATACTT TTTCAATCA ATTGCAITAA AAATAATTC	11700
AATTTTTCCG TCTTTATACG CTTTCTCTCC ATATAACCAC TTAGCCGAAT CTTCCCCACA	11760
TGCAAAACCA TGAGTTGCTA AGATTTTTAC CAAAATGTT ACTAATTTAC GCAATACTTT	11820
TTGAAAACCTG TTTTCTGTTA CATAAGCCAT ATGACTATGA ATAATCTAA TTTTACAACC	11880
AATTATTTTA GATAAGATCA GACCAATTGC AGATTTATAG CCATGGCAAT GAACTATATC	11940
ATAATCTCCT TTCTTTATTA TTCTAGCAAG AGAGAGAAAC TGATGTAGAG GCTTTTCTCT	12000
TAATAGAGGC ACATGATAAA CCTTTGCACC CAATCTTTT ATTTATCCT CTAATAATCC	12060
TTGTTCTTTT CCAGGCACAA TAAATCAAA TTGAATTTTT TTTCTATCAA TGTGAGAATA	12120
ATAGTTGAAT AGAAAACCTT CTACTCCACC ACTATCTAGT GTTGTAATA GATGTAATAC	12180
TTTAATCAAT GTTCTTCCTT AAGCTTAAGA TTCGCTTCTC TAATCTATT TCTGTTTTTT	12240
GTTTTCTAA ACTAATCTG TCCATGAAGT TATCACAATT CTTAATTAGC TGTTCCTGT	12300
CAAGGTTTTG AATATACAAA GCCAAACAAT CTTTTCCGA TTCATCCTTC ATAGTAAAA	12360
CGAAACCAAA ACCATTCTCT ATTGACACTT TTTCCATATA AGTATCTTCA CAAACTAAAA	12420
TAGGTTTATA CAACAATGCA GCAAAGTAGA GTTTATTAGA CAAAGCATAG TCTAGTAAGG	12480
GAGTGTGATT CCCGTATAAA TTCAAAACAA CATCTGTATT CTTATAAAAA GACATGGTAT	12540
CTTTAGGCTG GAATGTGTCC ACCAAGTTAA CATTGCTGAT ATTTTTTCT TGACAAAATT	12600
CCCTTAATTC TCCTGCATTA GTACCTATAA AATTCAACTG AAATCGACTG TCATTGCAA	12660
AAAAATCGAT TATTTTTTTA TTTTGTCTT GAAAACGAAT TAAACCAATG TAGGAAAGTT	12720
GAATTGGAAA CGTACTATTA TTTTTTAACT GCTTTACCTC GTTTAATTCT ATCATATTGG	12780

355

GTAGGTTATG GGTAGTAAAA TACTCTCCCA TTGGTAAAAA AAATTTATAG CCGTCTGAAG	12840
AAACGATATT CATTAAAGAA TTTTTCACCA ATTGTTTCTG AACCAAACGA TAAACCAAAA	12900
ATTTTTCATA ACTGTAATCA CGAATATCAT AAATATATCT ATTTTAAAT GAAAAGAGAA	12960
GAAAATCTAC TAAAATGAAA GACACAATAC TATGTAACGG CAATATCATA TCATAATCAT	13020
TTTCTTTTAG CTCTTTTTA ATTTCTTTT TGAATTTTAC ATAACCTAAT ATCTTACTTA	13080
ATTTTCCTTT ACCAGAAAAA GAAATACGAT AGTAGTTTGG TTTTGTAATA ATCTCGTTAA	13140
TATCTTATC CCAATATATA ACATCGTAAC TAATAGACAG TTTCTTCAAT AATTCTTTAT	13200
AAAAATGAA GTAAGGAGTT AGATATATAT TATCAGATAG TATAAACAGT ACTCTCATTA	13260
AATTATCTT TCTTACTTTC CCTCTCTAAA CATGTCTCCA GTTCGAGCAT AAATGCTCT	13320
TTTGAAAAGT GATTTTCATA GTAACAACGA GCTTCTTTT CTAACCTCTT TTGTCTCTTA	13380
ATAGATAACA TACTAAATTT ACAAATATTT TTTGCCAATT GTTTTACATC TCGTTGCGGA	13440
CTAACATATC CACAATTTGC TTCTTCTACA ATTATTTTAG CATCTCCTGA AATTGCACCT	13500
ATAATTGTT TGCCTGCCGC CATATAAGAK TGTACCTTCC CAGGTATAGT ACGAGAACT	13560
ATCGAGTCTC CTATTAAAGA AACTAACATA GCATCTGATT TTTTATAGAA GGATGGCATT	13620
TCCTCCAAAG AACGTCTTCC ATAGAAGGAA ATATTCTTTA ACTCCAATTC ATGAGCTAAT	13680
GCTTTCATGC TTAACAATTC CGTACCATCT CCAACAAAAT GAAAATGAAT TTTCTTGGGT	13740
AAATTGGTAT TCTTCTCTAT CAAACTGGCA GCTTTCAAAA TAGTTTCCAA ATTTTGTGCT	13800
TTGCCAATAT TACCAGCAAA AGTTAGGTCA ACACTTTCTT TATTAATAT AGATTCATCA	13860
GGGATAAAAA GATCTTCTGC ATATTGTGGC AAATATGTAA TCTTTTCTTC GGATATGTCA	13920
AATTGCTTCA CAAAATAATT TTTAAATGAT CGACTAGTGA CAAATATATA ATCACTAGCT	13980
CGGTAAACTT TTTTGTAGAT AAATTTAAAC AGCTTGAAAA TCAAGCCATC TTGTTTCACT	14040
CCACCTACGG TTAAACTATC TGGCCAAACA TCCATACAAT ATAGAAACAT CGGTTTCTTA	14100
TATTTTTTTT TATAAGCCAT ACCAGCCCAT GCCATCATAA CTGGAGACAA TTGGTTAACC	14160
AATACACAGT CAAAATTCGA TCCATCTTTC GTTTTATACC TCCCAATAA AACTCCTAAA	14220
GTAGAACTAA TTGCAAAGCT AAAATAATTC AACAATCGAA ATACAACACT TTTTTTCTA	14280
GGGATTGTAT AAGAACGATA TATCGTAACA CCTTCTATAA TCTCAGTCT TTTTATTATA	14340
TGACGATAAT CTGCATATAT CTTCCTTCA GGGTAATTAG GAATCCCAGC CAAAACAGAG	14400
ACTTCATGCC CTTTTCGAAC TAAATCTTCA CAAATATCTG ACAACCTGAA TGGTCTGGC	14460
TTATAATGTT GGCAAACAAA TAGTATTTTC ATGTGCCAAT TTAACCTTCT TTCTTACCAC	14520

356

TACCCCTCTAC AATACCTTTT CGTTTCAGTA CGTAAGGTAT TGTCTTAACT ATACATCTAA	14580
TATCCATTAT CAAAGACAGA TGTTTAACAT AGTAGCCATC TAACTCCGTC TTCATCTCAA	14640
CAGACAAAGT ATCACGCCCG TTAATTTGTG CCCATCCAGT TAACCCTGGC AAGATATCAT	14700
TTGCTCCATA CTTATCTCTC TCTGCAATCA AATCTAGTTC ATTTATACCC GCTGGTCTAG	14760
GACCTACAAT ACTCATATTA CCAACAAGAA TATTAAACAA TTGTGGTAGT TCATCCAAAG	14820
ATGTTTTTCG CAAGAAAGCC CCTACTTTTG TAATCyATTG CTCTGGATTA TATAAGTTTC	14880
GAGGCGCCAC ATTTTtagGT GCATCTATTT TCATAGACCT AAATTTCAA ATATAGAAGT	14940
ATTCCTTATG AATACCAAAG CGTTTTTGCT TAAATATAAC CGGACCTTCT GAATCAAGTT	15000
TAATCGCAAT TGCAATTATC ATAAAAACCG GACACAATAT TATTATCCCT ATTAAAGATA	15060
ATAATATATC ACCTAATCGT TTTATTATAC CGTACATAAA CAACCTCCAA CTATAAATTC	15120
TATTTCCATT TTTCACTCTA TTTCCATTTG ACAAATTAAT TCAGGCAGTA CATGCAACTA	15180
CAGAAACTCA ATATATATTT GGTCACTCAA TGATTTTCAG AAATATAATT CTTTATCCT	15240
CTACGTCAGA TAAAACTTTT CTCCATCTAA ACAAATTTA TTTGTTTCAG TAATATATGA	15300
GTTCTCAATA ATGAATTAGA AGGTCCAGTT CAATTATTCT TCCAAATAGA CCGAATATTA	15360
TTTGAAGACA TATCGGTTTC TGAAATTGCA ATCAGTACAT AAGCTAATAA ACTGATAAGT	15420
ATGCTCTGTA AGAATGCCAG AGTTATATTG TAGTCCCCCT CCATACTATA TTCATTTTAT	15480
TTTTTACCAT AATTTCCATA GGAACCGTAA ACTCCATACT TATTAACCGA GATATCCAAT	15540
TTATTTAAAA CAACTCCTAG GAACAGTTTC CCTGTTTGT TTAATTGTTG TTTTCGTTTT	15600
TGGATATCAC GTTTATTTCG CTCACCTGTT GCTGTTACCA AGATGGACGC ATCACACTTT	15660
TGAGTGATAA TTGCCGCATC AATAACAATT CCAATAGGCG GTGTATCAAT AATGATATAA	15720
TCAAAATATT TACGCAATGT TTCAATCATA TCATTAAAAAT TTTTACTTTG TAACAAGGCT	15780
GTAGGGTTTG GTGATACAGA TCCCGATTGA ACTACAAATA AATTTCAAT ATTTGTATCA	15840
CATAAACCGT GAGATAAATC AGCTGTCCCA GATAAAAAAT CTGTTAGCCC TGTAATTTT	15900
TCACGAGATT TAAAACTCC TAACATAACT GAATTTGAG TATCGCCATC GATCAAAAGA	15960
GTTTTATAGC CTGCACGCGC AAACGACCAT GCTATATTTA TGGAAGTAGT TGTTTTTCCT	16020
TCCCCAGGGT TAACAGAAGT AACGGAAATT ACTTTTAGTT TATCTCCGCT CAACTGTATA	16080
TTTGTACACA AGGCATTGTA ATATTCTTCT GCCTTCTTAA TGAACCTCAG TTTTTTTTGT	16140
GCTATTTCTA ATGTCGGCAT CCTTCTCTCC TATTTCAACT TACCCAAGTT TGGCACAAC	16200
CCCAAAAGTG TCATCTGCAA TGTATTTTCG ATATCTTCCG GACGTTTCAC ACGAGTATCC	16260
AAAAGTTCAA GATGAAGAAC TATAACACTA GTTCCAATCA CCCCTGCCAA AAAACCAATT	16320

357

AGTGTATTGC GTTTAATATT TGGCGAAGAC GGGGATATCG CCGGCCTTGC CTCCTCCAGT	16380
GTTGTACAGT CAGAAACACG AGTAATACTG ATAATTTTTT GAGCAGCTAC TTCTCTCAAA	16440
GAGTTAGCGA TACGGCTTGC CTCTTCAGGA ACTCGATCAT TAACTGAAAT AGAGACAATA	16500
CGGGTATCAA CTGGTACTGT CACTTTAATT TTATTAGCCA AACCTTTTGG CGTCAAATCT	16560
AGTTTCAAAT CAGAAACAAC TTCCTCCAAA ACATCCTGCG AAAGGATAAT CTCACGGTAG	16620
TCTTTTACCA GATAAGTTCC TGCCTGCAAA TCCTGATTG TCAACCCCGG CTTGTCTCCT	16680
TGATTGCGAT TCACTACGTA AATTGCGGTG GTACTCGTAT ATTCTGGCTT AACAAATAAA	16740
GTGCTATATG CAAAAGCCCC CGCACCTGTC ACAAGTGCCA CTATTAAAAT CATTAGCTTG	16800
CGTTTCCACA AGCTTTTAAC TAATTGAAAT ACATCGATTT CTATCGTATT TTGTCTTTTC	16860
ATCATTTCTC CTAAATTAGT TGATCCATTA CAATTTTTCG AGGATTGTCT ATAAAAAGTT	16920
CCTGAGCCTT CGCTTCTCCG TATTTTGGG TAACAAGGTC ATATGCTTCT GCCATATGAG	16980
GAGGTCTACC GTCTAGATTG TGCATATCAC TTGCAATGAC ATGAACCAAA TCCTGCTCTA	17040
AAAAATACTG AGCTCTTTT TTCATGAATT TATAACGTTT GCCAAAAAGT TTGGGTTTGA	17100
GGACATGTGA ACTATTTACT TGGGTGTAAC AGCCCATATC GATCAGTTCT CGAACGCGTT	17160
TTTCATTATT TTCAAGAGCA TCATAGCGCT CAATGTGGGC AATGACTGGA GTAATTCCCA	17220
ACATCAAGAT CTTGCTCAAG GCGCTATGAA TATCGCGATA AGGAGTGTTT AACTAAACT	17280
CTATCAAGGC ATAACGACTA TCATTGAGGG TCGGAATCCG CTTTTTTTCC ACCTTATCCA	17340
GAACATCTGG TGTGTAATAA ATTTAGCCCC CGTAAGCAAT GACCAAGTCA CTCGCCACTT	17400
CCTTAGCTAT TTCCCGAACC TGAAGAACT TTTCTGCTAT CTTCTCTTCC GGAGTTTCAA	17460
ACATGCCCTT GCGACGGTGA GAGGTAGAAA CAATGGTTCG CACCCCTGT CTGTAGGATT	17520
CTGCCAAGAG AGCCTTGCTT TCCTCTCTTG ACTTGGGACC GTCATCTACA TCAAAAACGA	17580
TATGCGAATG GATGTCTATC ATTTATCTA CCCTCCATCA CATCCTGTAT AGCTGCTTTA	17640
ACTACAGCTA AACTACTATC ATCTATTTCC ATCACATAGA GGTTACTGTC TGGCATTGCA	17700
TAAGAAGGAA GATCCATCCG ACCTGTCCCT TTTAAATCTT GAGAATTTAC TTTATAATTC	17760
CCTCCACTTT CTAAGTACG ATTGACCAAA TTTATCATGG TCTCAAGTGG CATATTTGTT	17820
TGGATAGAAT CTTGCAAGCT ATTAATGATC GTACTATAAT TTTTCAGCAC TTCGGTTGAC	17880
GTAAATTTTT GAAGGATAGC CACAATCACC TTTTGTGAT GCGCCCCGCG GTCACGATCG	17940
CCATCTGCTA GGGAGTAGCG CTCACGAACA AAACCGAGAG CCTGTTCTGA ATCAAGATGA	18000
ACATTGCCTG CAGGGTAATA CTTTCCATTC GTATGGGCAG TAAATTCTTG ATCATTATAA	18060

158

ACATCAATTC CACCCAACAA ATCAATCAAT TTCAAAAACG AAGTGAAGTT CAATCGCACA	18120
TAGTAATTGA TATCCACTCC ATAGAGATTT TCTAAGGTGT GAATGGACGA ATCAACTCCA	18180
TAAATGCCCG CATGAGTCAA TTTATCTTTT TGATTATTTC CACCATCTGC GATTGGTACA	18240
TAGGCATCAC GTGGCGTTGT GGTCAAGAGG ATTTCTTGG TATCTCGATT GACAGTCATC	18300
AGGATGTTGA CATCTGATCG CGACACCGAA CTAATAGGAC CATAGGTGTC AATTCCACTA	18360
ACATAGATAT TGAAAGACTG ACTCTTAGAC GTCTTAGGAG CTTCTACTTT TTTAGTGAAT	18420
CCCTTAGTAT AAATCTTTTT TATCTTCGAT GCGTAGTCTG GATACTCTGA CTCGATGATG	18480
TTTTCAAAGA CACTATTAG GACAATGGCC TTAGTCTCCC CTGCAATCAA ACTCTGTAA	18540
GCTGCCAAGT AAGACGAACT CTGGTTGACC GTCAAATCGG TATTCTGACT TGACTTGATA	18600
TCAGCTAGTA ATTTCTGAAT ATTTTCATTA TTAGTCCCAG TCGGTGCTGT CACACTCGTC	18660
AGTTGCGTAA CATTTTCGAT CTCACTATCT GCTAAAACAG CGACACTGAT TGAATATTCT	18720
GAGTAATTAG AAGTCGCATT TAAACGATTG GTCAGTCCAA CAAACTGCTG TACTGCAAAG	18780
AGCGACACAG AGCTGACAAG GATAGAGAAC ACCAACAGAA AAATAGTAAA CTTTTCAGCT	18840
TTTTTATAGA TAATCAAGAG TAGCCCTACC AAGGCAACTA GTAGGACTAA CGCAGTTACC	18900
ACTAGATTAA GATATCTAAA AGCAAGGATA TTCTACTTAA AGATTAAGAA CAATAAAAAA	18960
CAAACTAACA ATAAATAAAT AGTCAGCAAA ACTATATTAA CACTTCGCTT CACTTTCTGT	19020
GAACGTGATT TTTTAAAACG TCTACTCATG ATTAATACCT ATACATTGAA CATTATACGA	19080
TTATATCACT TTTTACGGT AATGTCTACA CCTTTATTTT TACTATCTGC ATCTTTAAGT	19140
ATCTTAGTAG ACTTCCCGCG AAACAAAAAT ATAGTAAAAT GAAATAAGAA CAGAACAAAT	19200
CGTTCAGGAC AGTCAAATCG ATTTCTAACA ATGTTTTAGA AGCAGAGGTG	19250

## (2) INFORMATION FOR SEQ ID NO: 36:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21706 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AAAGTTGAAA GACTGCTAGC TGTTTTGTAT ACCAATCGTT TCCAACTACA GAGCAAACAG	60
TATACAAAGT TTGTTTTTGG ATGTAAGCTT CTTGATGGAC AATCCAAGA AAATCAAGAA	120
ATTGCTGACC TTCAATTTTT TGCCATTGAC CAACTGCCGA ACTTATCTGA AAAACGCATT	180
ACCAAGGAGC AAATAGAGCT TCTTTGCCAG GTTTATCAAG GTCATAGGGG GCAATATCTT	240

GACTAAGAAG ATGATTATCG TATTTCTAAA TCCATTTTTA ACAACTAGCA TGGTATAATA	300
ATATGCAGGA AAATTTTGAA TTATGAGGAA GACTAGATGA ATTTATGGGA TATTTCTTTT	360
ACGACTCAGG CAACCGAGCC GCCCAAATTT GACCTTTTTT GGTATGTTAG CCTATTTACG	420
CTCTTAGCCT TAACCTTTTA TACAGCCCAT CGCTATCGTG AAAAGAAGGT TTACCAACGA	480
TTTTTC AAA TCTTGCAGAC TGTTCACTTA ATCCTTCTTT ATGGTTGGTA CTGGGTCAAT	540
CATATGCCAC TGTCAGAAAG CCTACCTTTT TACCATTGCC GTATGGCTAT GTTTGTGGTA	600
CTCTTGCTTC CTGGTCAATC CAAATATAAA CAATACTTTG CATTATTGGG AACATTTGGG	660
ACATTAGCAG CCTTTGTTTA TCCAGTGCCA GATGCTTACC CTTTCCACA TATCACCATT	720
CTATCCTTTA TCTTTGGTCA TTTAGCACTC TTGGGGAAC CTCTAGTTTA TCTATTGAGA	780
CAGTATAATG CGCGATTGCT GGATGTGAAG GGAATTTTTC TCATGACCTT TGGCCTAAAT	840
GCCTTGATTT TTGTGGTCAA TTTGGTGACA GGTGGCGATT ACGGATTTTT GACAAAACCG	900
CCATTGGTTG GGGATCACGG TCTAGTAGCT AATTATTTAC TTGTTTCAAT TGTGCTGGTA	960
GCTACTATCA GTTTGACTAA GAAAATCTTA GAATCTTTT TAGCTCAAGA AGCAGAAAAA	1020
ATGATTGCAA AGGAAGCTTA ACACAGAGCT TTCTTTTTTG CTCTTAGAGA GTTTTACAA	1080
GCAGCTTATA AAATAAGAAT TTCTGAATAG ACAAACCTCAA AAAATGGCTG GGAATTTAG	1140
GAAAAAAGCA AGCACGATTA AATTTTTTGT GTTATAATAT TTTGTGAATA GCTATGCCTA	1200
TGTTTAGCTA TGGAATAATA CGAAGTGCGA AACTTGGAAG ATAGAGAGGA AGCGATGTAA	1260
TGGCTAGAGA AGGCTTTTTT ACAGGTCTAG ATATTGGAAC AAGCTCTGTC AAGGTGCTTG	1320
TGGCCGAGCA GAGAAATGGT GAATTAAATC TAATTGGCGT GAGTAATGCC AAAAGTAAAG	1380
GTGTAAAGGA TGGAATTATT GTTGATATTG ATGCAGCAGC AACTGCTATC AAGTCAGCCA	1440
TTTCCCAAGC GGAAGAAAAG GCAGGCATTT CGATTAAATC AGTGAATGTC GGCTTGCCCTG	1500
GTAATCTTTT GCAGGTAGAA CCAACTCAGG GGATGATTCC AGTAACATCT GATACTAAGG	1560
AAATTACGGA TCAAGATGTT GAAAATGTTG TCAATCAGC TTTGACAAAG AGTATGACAC	1620
CTGACCGTGA AGTCATTACC TTTATTCTG AAGAATTTAT TGTGGATGGT TTCCAACGGA	1680
TTCTGACCC ACGTGGCATG ATGGGGGTTT GCCTTGAAAT GCGTGGTTTG CTTTATACAG	1740
GACCTCGTAC TATCTTGAC AATTGCGTA AGACGGTTGA GCGTGCAGGT GTTCAGGTTG	1800
AAAATGTTAT CATTTACCA CTAGCAATGG TTCAGTCTGT TTTGAACGAA GGGGAACGTG	1860
AATTTGGTGC TACAGTGAAT GATATGGGGG CAGGTCAAAC GACTGTCGCT ACAATCCGTA	1920
ATCAAGAACT CCAGTTCACA CATATTCTCC AAGAAGGTGG AGATTATGTA ACTAAAGATA	1980

360

TCTCCAAGGT TTTGAAAACC TCTCGCAAAT TAGCGGAAGG CTTGAAACTG AATTACGGGG	2040
AAGCCTATCC GCCTCTTGCA AGCAAAGAAA CCTTCCAAGT AGAGGTTATT GGAGAAGTAG	2100
AAGCAGTCGA AGTGACGGAA GCCTACTTGT CAGAAATTAT TTCTGCACGA ATCAAGCACA	2160
TCCTTGAACA AATCAAGCAA GAATTAGATA GAAGGCGTCT ATTGGACCTC CCTGGTGGTA	2220
TTGTCTTAAT CGGTGGGAAT GCCATTTTAC CAGGTATGGT TGAGCTTGCT CAGGAAGTCT	2280
TTGGCGTCCG TGTCAAGCTT TATGTTCCAA ATCAAGTTGG TATCCGTAAT CCAGCCTTTG	2340
CGCATGTGAT TAGTTTATCA GAATTTGCGG GTCAATTAAC AGAAGTTAAT CTTTGTGGCTC	2400
AGGGAGCGAT AAAAGGTGAG AATGACTTAA GTCATCAGCC AATTAGTTTT GGTGGGATGC	2460
TGCAAAAAAC AGCTCAGTTT GTACAATCAA CGCCTGTTCA ACCAGCTCCT GCTCCAGAAG	2520
TAGAGCCGGT GCGGCCTACA GAACCAATGG CGGATTTCCA ACAAGCTTCA CAAAATAAAC	2580
CGAAATTAGC AGATCGTTTC CGTGGATTGA TCGGAAGCAT GTTTGACGAA TAAAGAGGAA	2640
AAATAAATTA TGACATTTTC ATTTGATACA GCTGCTGCTC AAGGGGCAGT GATTAAAGTA	2700
ATTGCTGTCG GTGGAGGTGG TGGCAATGCC ATCAACCGTA TGGTCGACGA AGGTGTTACA	2760
GCGGTAGAAT TTATCGCAGC AAACACAGAT GTACAAGCAT TGAGTAGTAC AAAAGCTGAG	2820
ACTGTTATTC AGTTGGGACC TAAATTGACT CGTGGTTTGG GTGCAGGAGG TCAACCTGAG	2880
GTTGGTCGTA AAGCCGCTGA AGAAAGCGAA GAAACACTGA CGGAAGCTAT TAGTGGTGCC	2940
GATATGGTCT TCATCACTGC TGGTATGGGA GGAGGCTCTG GAACTGGAGC TGCTCCTGTT	3000
ATTGCTCGTA TCGCCAAAGA TTTAGGTGCG CTTACAGTTG GTGTTGTAAC ACGTCCCTTT	3060
GGTTTTGAAG GAAGTAAGCG TGGACAATTT GCTGTAGAAG GAATCAATCA ACTTCGTGAG	3120
CATGTAGACA CTCTATTGAT TATCTCAAAC AACAAATTGC TTGAAATTGT TGATAAGAAA	3180
ACACCGCTTT TGGAGGCTCT TAGCGAAGCG GATAACGTTT TCCGTCAAGG TGTTCAGGG	3240
ATTACCGATT TGATTACCAA TCCAGGATTG ATTAACCTTG ACTTTGCCGA TGTGAAAACG	3300
GTAATGGCAA ACAAAGGGAA TGCTCTTATG GGTATTGGTA TCGGTAGTGG AGAAGAACCT	3360
GTGGTAGAAG CCGCACGTAA GGCAATCTAT TCACCACTTC TTGAAACAAC TATTGACGGT	3420
GCTGAGGATG TTATCGTCAA CGTACTGGT GGTCTTGACT TAACCTTGAT TGAGGCAGAA	3480
GAGGCTTCAC AAATTGTGAA CCAGGCAGCA GGTCAAGGAG TGAACATCTG GCTCGGTACT	3540
TCAATTGATG AAAGTATGCG TGATGAAATT CGTGTAACAG TTGTTGCAAC GGGTGTTCGT	3600
CAAGACCGCG TAGAAAAGGT TGTGGCTCCA CAAGCTAGAT CTGCTACTAA CTACCGTGAG	3660
ACAGTGAAAC CAGCTCATTC ACATGGCTTT GATCGTCATT TTGATATGGC AGAAACAGTT	3720
GAATTGCCAA AACAAAATCC ACGTCGTTTG GAACCAACTC AGGCATCTGC TTTTGGTGAT	3780

361

TGGGATCTTC	CCCCGGAATC	GATTGTTTCGT	ACAACAGATT	CAGTCGTTTC	TCCAGTCGAG	3840
CGCTTTGAAG	CCCCAATTTC	ACAAGATGAA	GATGAATTGG	ATACACCTCC	ATTTTTCAAA	3900
AATCGTTAAG	TAAATGAATG	TAAAAGAAAA	TACAGAACTT	GTTTTTCGAG	AAGTTGCAGA	3960
GGCTAGTCTG	AGTGCTCATC	GAGAGAGTGG	TTCGGTCTCT	GTCATTGCAG	TTACCAAGTA	4020
TGTAGATGTA	CCGACAGCGG	AAGCCTTGCT	TCCGCTAGGT	GTCCATCATA	TCCGTGAAAA	4080
TCGTGTAGAT	AAGTTTCTGG	AAAAATATGA	AGCTTTAAAA	GATCGAGATG	TGACTTGGCA	4140
TTTGATTGGT	ACCTTGCAAA	GACGTAAGGT	GAAAGATGTC	ATTCAATACG	TTGATTATTT	4200
CCATGCATTG	GACTCAGTAA	AGCTAGCAGG	GGAAATTCAA	AAAAGAAGTG	ACCGAGTCAT	4260
CAAGTGTTC	CTTCAAGTAA	ATATTTCTAA	AGAAGAAAGC	AAACACGGTT	TTTCGAGAGA	4320
GGAAGTCTG	GAAATCTTGC	CAGAGTTAGC	CAGACTAGAT	AAGATTGAAT	ATGTTGGTTT	4380
AATGACGATG	GCACCTTTTG	AGGCTAGCAG	TGAGCAGTTG	AAAGAGATTT	TCAAGGCGGC	4440
CCAAGATTTA	CAAAGAGAAA	TTCAAGAGAA	ACAAATTCCA	AATATGCCTA	TGACCGAGTT	4500
AAGTATGGGA	ATGAGTCGTG	ATTATAAAGA	AGCGATTCAA	TTCGGTTCCA	CTTTTGTTCC	4560
TATAGGTACA	TCATTTTTTA	AGTAGGAGAG	AACCATGTCT	TAAAAGATA	GATTTCGATAG	4620
ATTTATAGAT	TATTTTACGG	AGGATGAGGA	TTCAAGTCTC	CCTTATGAAA	AAAGAGATGA	4680
GCCTGTGTTT	ACTTCAGTAA	ATTCTTCACA	GGAACCGGCT	CTCCAATGA	ATCAACCTTC	4740
ACAGTCGGCT	GGCACAAAAG	AGAACAATAT	CACCAGACTT	CATGCAAGAC	AACAGGAATT	4800
GGCAAATCAG	AGTCAGCGTG	CAACGGATAA	GGTCATTATA	GATGTTTCGT	ATCCTAGAAA	4860
ATATGAGGAT	GCAACAGAAA	TTGTTGATTT	ATTGGCAGGA	AACGAAAGTA	TCTTGATTGA	4920
TTTTCACTAT	ATGACAGAGG	TCCAGGCTCG	TCGTTGTTTG	GACTATTTGG	ATGGAGCTTG	4980
TCATGTTTTA	GCTGGAAATT	TGAAAAAGGT	AGCTTCTACC	ATGTATTTGT	TGACACCAGT	5040
GAACGTTATT	GTAAATGTTG	AAGATATCCG	TTTACCAGAT	GAAGATCAAC	AGGGTGAGTT	5100
CGGTTTTGAT	ATGAAGCGAA	ATAGAGTACG	ATAATGATTT	TTTTAATTCT	TATGATTTAT	5160
AATGCAGTGG	ATATTTACTC	CCTGATTTTG	GTAGCCTTCG	CTGTCATGTC	TTGGTTTCCA	5220
GGTGCCTACG	AATCCAGTTT	AGGTCGTTGG	ATTGTAGCGT	TGGTGAAACC	AGTGCTTGCT	5280
CCCTTGCAAC	GCCTGCCTTT	ACAGATAGCG	GGTCTTGATT	TATCTGTTTG	GGTTGCGATT	5340
GTTTTGGTTC	GATTTTTTAGG	AGAAAACCTA	GTGCGTTTTT	TGCCGATGAT	AGGATGAATA	5400
AAGGGATTTA	TCAGCATTTT	TCCATAGAAG	ATCGTCCATT	TCTTGACAAG	GGAATGGAAT	5460
GGATAAAGAA	GGTAGAAGAT	AGCTATGCTC	CTTTTTTAAC	TCCTTTTATC	AATCCTCATC	5520



362

AGGAGAAGCT ATTAAAGATT TTGGCCAAAA CCTATGGTCT TGCTTGTAGC AGTAGTGGGG	5580
AATTCGTCTC GAGTGAGTAT GTTCGAGTTT TATTATACCC AGATTATTTT CAACCAGAGT	5640
TTTCAGATTT TGAAATATCT CTCCAGGAAA TTGTGTATTC CAATAAATTT GAACATTTAA	5700
CGCATGCTAA GATTTTAGGG ACAGTCATCA ATCAATTAGG GATTGAACGG AAACTTTTTG	5760
GAGATATCCT AGTAGATGAA GAACGGGCGC AGATTATGAT TAATCAGCAG TTTCTTCTTC	5820
TCTTTCAAGA TGGACTAAAG AAAATTGGTC GTATACCTGT TTCGCTGGAG GAACGTCCTT	5880
TCACCGAGAA AATAGATAAG CTAGAACAGT ATCCGAGAACT GGATTTATCT GTGTCTAGTT	5940
TTGCGATTAGA TGTTCTTTTA TCAAATGTTT TGAACTATC TAGGAATCAA GCAAACCACT	6000
TGATTGAAAA GAACTTGTCT CAAGTAAATT ATCATGTGGT AGACAAATCA GATTACACTG	6060
TTCAAGTTGG AGACTTGATT AGTGTGAGAA AATTTGGTCG CTTGAGATTA CTTCAAGATA	6120
AGGGACAAAC GAAAAAGAG AAGAAAAAAA TAACCGTCCA GTTATTATTA AGTAAGTGAG	6180
GAATAGAATG CCAATTACAT CATTAGAAAT AAAGGACAAG ACTTTTGGAA CTCGATTGAG	6240
AGGTTTTGAT CCAGAAGAAG TCGATGAATT TTTAGATATT GTGGTTCGTG ATTACGAAGA	6300
TCTTGTGCGT GCGAATCATG ATAAAAATTT GCGTATTAAG AGTTTAGAAG AGCGTTTGTC	6360
TTACTTTGAT GAAATAAAG ATTCATTGAG CCAGTCTGTA TTGATTGCTC AGGATACAGC	6420
TGAGAGAGTG AAACAGGCGG CGCATGAACG TTCAAACAAT ATCATTGCTC AAGCAGAGCA	6480
AGATGCGCAA CGCTTGTGG AAGAAGCTAA ATATAAGGCA AACGAGATTC TTCGTCAAGC	6540
AACTGATAAT GCTAAGAAAG TCGCTGTTGA AACAGAAGAA TTGAAGAACA AGAGCCGTGT	6600
CTTCCACCAA CGTCTCAAAT CTACAATTGA GAGTCAGTTG GCTATTGTTG AATCTTCAGA	6660
TTGGGAAGAT ATTCTCCGTC CAACAGCTAC TTATCTTCAA ACCAGTGATG AAGCCTTTAA	6720
AGAAGTGGTT AGCGAAGTAC TTGGAGAACC GATTCAGCT CCAATTGAAG AAGAACCAAT	6780
TGATATGACA CGTCAGTTCT CTCAAGCAGA AATGGCAGAA TTACAAGCTC GTATTGAGGT	6840
AGCCGATAAA GAATTGTCTG AATTTGAAGC TCAGATTAAA CAGGAAGTGG AAGCTCCAAC	6900
TCCTGTAGTG AGTCCTCAAG TTGAAGAAGA GCCTCTGCTC ATCCAGTTGG CCCAATGTAT	6960
GAAGAACCAG AAGTAGCTCC AATGCATCCG ATAGGTCCAA CACCAGCTAC AGAACTGTT	7020
GATTCAATAC CGGGATTGTA AGCACCGCAA GAATCTGTTA CAATTTTATA AGAAATATTC	7080
TGAGAACAAT ATCTTATCCT TATATTTCCA GCGAGCAGGA GATGGTGTGA GTCCTGTAAT	7140
CCCTATTGAT AAGATTATCC TCTCAAAAAC TCAAGTCTGA AGCTAGTAAG ATTTGACGTT	7200
TCCCACGTTA CGGGATAAGA GGGAGAAAGA CTAAATCTTT TTCCGAATAA AGGTGGTACC	7260
ACGATTTTCG TCCTTTTGG AAGTCGTGGT TTTTAAATTT TTATTATTTA TAAAGGAGAT	7320

363

ACCATGAAAC TCAAAGACAC CCTTAATCTT GGGAAAAC TG AATTCCCAAT GCGTGCAGGC	7380
CTTCCTACCA AAGAGCCAGT TTGGCAAAAG GAATGGGAAG ATGCAAAACT TTATCAACGT	7440
CGTCAAGAAT TGAACCAAGG AAAACCTCAT TTCACCTTGC ATGATGGCCC TCCATACGCT	7500
AACGGAAATA TCCACGTTGG ACATGCTATG AACAAGATTT CAAAAGATAT CATGTTCGT	7560
TCTAAGTCTA TGTCAGGATT TTACGCACCA TTTATTCCTG GTTGGGATAC TCATGGTCTG	7620
CCAATCGAGC AAGTCTTCTC AAAACAAGGT GTCAAACSTA AAGAAATGGA CTTGGTTGAG	7680
TACTTGAAAC TTGCGCTGA GTACGCTCTT TCTCAAGTAG ATAAACAACG TGAAGATTTT	7740
AAACGTTTGG GTGTTTCTGG TGA CTGGGAA AATCCATATG TGACCTTGAC TCCTGACTAT	7800
GAAGCAGCTC AAATTCGTGT ATTTGGTGAG ATGGCTAATA AGGGTTATAT CTACCGTGGT	7860
GCTAAGCCAG TTTACTGGTC ATGGTCATCT GAGTCAGCAC TTGCTGAAGC AGAGATTGAA	7920
TACCATGACT TGGTTTCAAC TTCCCTTTAC TATGCCAACA AGGTAAAAGA TGGCAAAGGA	7980
GTTCTAGATA CAGATACTTA TATCGTTGTC TGGACAACGA CTCCATTTAC CATCACAGCT	8040
TCTCGTGGTT TGACGGTTGG TGCAGATATT GATTACGTTT TGGTTCAACC TGCTGGTGAA	8100
GCTCGTAAGT TTGTCGTTGC TGCTGAATTA TTGACTAGCT TGTCTGAGAA ATTTGGCTGG	8160
GCTGATGTTT AAGTTTTGGA AACTTACCGT GGCCAAGAAC TCAACCACAT CGTAACAGAA	8220
CACCCATGGG ATACAGCTGT AGAAGAGTTG GTAATTCTTG GTGACCACGT TACGACTGAC	8280
TCTGGTACAG GTATTGTCCA TACAGCCCTT GGTTTTGGTG AGGACGATTA CAATGTTGGT	8340
ATTGCTAATA ATCTTGAAGT CGCAGTGACT GTTGATGAAC GTGGTATCAT GATGAAGAAT	8400
GCTGGTCTTG AATTTGAAGG TCAATTCTAT GAAAAGGTAG TTCCAACGT TATTGAAAAA	8460
CTTGGTAACC TCTTCTTTC CCAAGAAGAA ATCTCTCACT CATATCCATT TGA CTGGCGT	8520
ACTAAGAAAC CAATCATCTG GCGTGCAGTT CCACAATGGT TTGCCTCAGT TTCTAAATTC	8580
CGTCAAGAAA TCTTGGACGA AATTGAAAAA GTGAAATTCC ACTCAGAATG GGTAAAGTC	8640
CGTCTTTACA ATATGATCCG TGACCGTGGT GACTGGGTTA TCTCTCGTCA ACGTGCTTGG	8700
GGTGTTCAC TTCTATCTT CTACGCTGAA GATGGTACAG CTATCATGGT AGCTGAAACT	8760
ATTGAACACG TAGCTCAACT TTTTGAAGAA TATGGTTCAA GCATTTGGTG GGAACGTGAT	8820
GCCAAAGACC TCTTGCCAGA AGGATTTACT CATCCAGGT CACCAAACGG CGAGTTCAAA	8880
AAAGAAACTG ATATCATGGA CGTTTGGTTT GACTCAGGT CATCATGGAA TGGAGTGGTG	8940
GTAAACCGTC CTGAATTGAC TTACCCAGCC GACCTTTACC TAGAAGGTTT TGACCAATAC	9000
CGTGGTTGGT TTAAC TCATC ACTTATCACA TCTGTGCCA ACCATGGCGT AGCACCTTAC	9060

364

AAACAAATCT	TGTCACAAGG	TTTTGCCCTT	GATGGTAAAG	GTGAGAAGAT	GTCTAAATCT	9120
CTTGGAATA	CTATTGCTCC	AAGCGATGTT	GAAAAACAAT	TCGGTGCTGA	AATCTTGCGT	9180
CTCTGGGTAA	CAAGTGTTGA	CTCAAGCAAT	GACGTGCGTA	TCTCTATGGA	TATCTTGAGC	9240
CAAGTTTCTG	AAACTTACCG	TAAGATTTCGT	AACACTCTTC	GTTTCTTGAT	TGCCAATACA	9300
TCTGACTTTA	ACCCAGCTCA	AGATACAGTC	GCTTACGATG	AGCTTCGTTT	AGTTGATAAG	9360
TACATGACGA	TTCCGTTTAA	CCAGCTTGTC	AAGACCATTTC	GTGATGCCCTA	TGCAGACTTT	9420
GAATTCTTGA	CGATCTACAA	GGCCTTGGTG	AACTTTATCA	ACGTTGACTT	GTCAGCCTTC	9480
TACCTTGATT	TTGCCAAAGA	TGTTGTTTAC	ATTGAAGGTG	CCAAATCACT	GGAACGCCGT	9540
CAAATGCAGA	CTGTCTCTTA	TGACATTCTT	GTCAAAATCA	CCAAACTCTT	GACACCAATC	9600
CTTCCTCACA	CTGCGGAAGA	AATCTGGTCA	TATCTTGAGT	TTGAAACAGA	AGACTTCGTC	9660
CAATTGTCAG	AATTACCAGA	AGTTCAAACCT	TTTGCTAACC	AAGAAGAAAT	CTTGGATACA	9720
TGGGCAGCCT	TCATGGACTT	TCGTGGACAA	GCACAAAAAG	CCTTGGAAGA	AGCTCGTAAT	9780
GCAAAAGTTA	TCGGTAAATC	ACTTGAAGCA	CACTTGACAG	TTTATCCAAA	TGAAGTTGTG	9840
AAAACCTAC	TCGAAGCAGT	AAACAGCAAT	GTAGCACAAAC	TTTTGATCGT	GTCTGAGTTG	9900
ACCATCGCAG	AAGGACCAGC	TCCGGAAGCT	GCCCTTAGCT	TCGAAGATGT	AGCCTTCACA	9960
GTTGAACGTG	CTACTGGTGA	AGTATGTGAC	CGTTGCCGTC	GTATCGACCC	AACAACAGCA	10020
GAACGCAGCT	ACCAGGCAGT	TATCTGTGAC	CACTGTGCAA	GCATCGTAGA	AGAAAACTTT	10080
GCGGAAGCAG	TCGCAGAAGG	ATTGAAGAG	AAATAAGATT	GAAAAGTCTA	GGCAAAATTC	10140
AATTTGAGAA	GAAAAGACAA	CTAATTTTAT	AGTCTATTAA	ACGCATTGTA	TCAGCTTTTT	10200
GAATACCTGA	TATGATGCGT	TTTTTATTTA	TTTTAAAAAT	TTGCGAGGTA	TGACTTTTTTA	10260
TACTCAACAA	GAATCAAAGA	GAACTTAGC	AAGCTAACAG	TAGTAAGATA	AAATAGGAAT	10320
TTGATATTAG	GGATAAGATT	GGTAAATAGT	GTAATATTTT	TACAACAATA	AATTTATATA	10380
GTTATTTCTG	GTTTCTGAAA	AGTATTATAT	TTTATTTTCAT	ATTATACAAA	TTTTTATTTT	10440
ATAATATCAG	AACATACTTT	TTTTAAAAGC	AAATATGATA	CAATTTTATT	TGAAAAAAAT	10500
AAAAAAGGAG	ATTTTATTAT	AAAATTAATA	AGACTTGCTT	TAATTAGTGG	TATCGTCCGT	10560
CTTGTGGGAG	GAATTTTACT	TCTTATTGGT	CCTTTTGTCT	TGTTGGGAAT	AGCGGTAAAC	10620
ACAGCTGCTA	CAACTCTTAA	TGGAGGAGCT	ACTGCAGGGG	CTTTTTCAGG	TGTAGCCTTA	10680
CTCTTGAATG	CCTTGAAGAT	TGCAATCTT	GTTCTTGCTA	TCATTGCTAT	TGTTTACTAT	10740
AAAGGAGATA	AGCGTGTAGG	TGCAGCTCCG	TCTGTACTAA	TGATTGTTTC	TGGTGGAGTT	10800
AGTCTCATTC	TATTCCGTTT	TTAGGATGGG	TTGGGGGGAT	TTTTGCTATT	ATCGGAGGAT	10860

365

CTCTATTCCT	TTCAACATTG	AAGAAATTCA	AATCAGAAGA	ATAAAAGGTA	TTTTAGCATG	10920
AAAAGAACAA	AAAAGTTTAT	CGGTATAGGA	GTAGCTCTAT	TATCTCTTTC	TCTTCTAGTT	10980
GCATGTGGAA	CATAAAGTTC	AAAGAATACT	TCAACAAGTA	ATGATGAGAA	GACAGTAGCA	11040
ACATCCAATA	GTTCAAAAGA	AACAATCACT	TTCGATACAC	CGGTTGTAAC	AGACGATGCC	11100
ATTGAATCAA	TACGCACTTA	TGCAGATTAT	ATAGATCTTT	ATAAAAATAT	TTTTGATGAT	11160
TATTTTACTA	AAGCTGAGGA	AGGTTTCAAA	GGCATAGCTA	TGGAAAATAA	TGACTCGTTT	11220
ACTAAACTAA	AAGAGTCAAC	TCAAAAATTA	TTCGATGCCG	AGAAAAAAG	GTTAAATAAT	11280
GAAGATAGAA	TAGAAACAAC	CAAAAACAAT	GTGATTGCCA	AACATTGTCA	AACAGTCCTT	11340
TCCTTTTGG	TTTGTAGTAG	CTTTTTTGTG	AAAAATTGTG	TAAAATAGAA	TAGATAAACG	11400
AGGGGAAACC	TCGGAAAATT	TAAAGGAGAA	TCCATCTAAT	GGTAAAATTG	GTTTTTGCTC	11460
GCCACGGTGA	GTCTGAATGG	AACAAAGCTA	ACCTTTTTCAC	TGGTTGGGCT	GATGTTGATT	11520
TGTCTGAAAA	AGGTACACAA	CAAGCGATTG	ACGCTGGTAA	ATTGATCAAA	GAAGCTGGTA	11580
TCGAATTTGA	CCAAGCTTAC	ACTTCAGTAT	TGAAACGTGC	TATCAAAACA	ACTAACTTGG	11640
CTCTTGAAAGC	TTCTGACCAA	TTGTGGGTTC	CAGTTGAAAA	ATCATGGCGC	TTGAACGAAC	11700
GTCACTACGG	TGGTTTGAAT	GGTAAAAACA	AAGCTGAAGC	TGCTGAACAA	TTTGGTGATG	11760
AGCAAGTTCA	CATCTGGCGT	CGTTCATACG	ATGTATTGCC	TCCAAACATG	GACCGTGATG	11820
ATGAGCACTC	AGCTCACACA	GACCGTCGTT	ACGCTTCACT	TGACGACTCA	GTTATCCCAG	11880
ATGCTGAAAA	CTTGAAAGTG	ACTTTGGAAC	GTGCTCTTCC	ATTCTGGGAA	GATAAAATCG	11940
CTCCAGCTCT	TAAAGATGGT	AAAAACGTAT	TCGTAGGAGC	TCACGGTAAC	TCAATCCGTG	12000
CCCTTGTAAG	ACACATCAAA	GGTTTGTGAG	ATGACGAGAT	CATGGACGTG	GAAATCCCTA	12060
ACTTCCCACC	ATTGGTATTC	GAATTCGACG	AAAAATTGAA	CGTCGTTTCT	GAATACTACC	12120
TTGGAAAATA	AAAAATTGTA	AGTCTAGAAT	TGATTTCTAG	GCTTTTATG	TTAGTATGGA	12180
AGTATGATAA	GGAATAAAAA	ACAAGATTAT	GTACTGGCCT	ACAAGCAACC	AGCTTCAACC	12240
ACTTACATGG	GTTGGAAGA	AGAAGCTTTA	CCGATAGGCA	ATGGTCTTTT	AGGAGCAAAA	12300
GTATTTGGCC	TTATAGGGCC	TGAACGGATT	CAATTTAATG	AAAAAAGTCT	CTGGTCTGGA	12360
GGTCCACTTC	CTGATAGTTC	AGATTATCAG	GGTGGAAATC	TTCAGGATCA	GTATGTTTTT	12420
TTAGCTGAGA	TTCCGGCAGG	TTTGGAGAAG	AGAGATTACA	ATCTGGCTAA	GGAAGTGGCT	12480
GAGCAGCACC	TAATTGGGCC	AAAAACGAGT	CAATATGGGA	CCTATCTGTC	TTTTGGGGAT	12540
ATTCACATTG	AGTTCAGCCA	GCAAGGTACG	ACTTTGTCTC	AGGTGACCGA	CTATCAGAGA	12600

366

CAGCTGAATA	TTAGTAAGGC	ACTTGCGACG	ACTTCTTATG	TCTATAAGGG	AACGCGATTT	12660
GAACGTAAAG	CTTTTGGCAG	TTTTCCAGAT	GATCTCTTGG	TTCAATGTTT	TACTAAGGAA	12720
GGGTTGGAAA	CTCTAGATTT	TACTATAGAA	CTATCCTTGA	CCTGTGATTT	GGCTTCTGAT	12780
GGAAAGTATG	AGCAGGAAAA	ATCTGATTAC	AAGGAGTGTA	AGTTGGATAT	TACTGATTCT	12840
CATATCTTGA	TGAAGGGAAG	AGTTAAGGAT	AATGATCTGC	GGTTTGCTAG	TTATCTAGCT	12900
TGGGAAACGG	ATGGAGATAT	TAGAGTTTGG	TCAGATAGGG	TTACATATC	AGGAGCCAGT	12960
TATGCCAATC	TCTTCTTGGC	CGCTAAGACG	GATTTTGCCC	AAAATCCTGC	TAGCAATTAT	13020
CGCAAGAAAC	TAGATTTAGA	GCAACAGGTG	ATAGACTTGG	TGGACACAGC	TAAAGAAAAG	13080
GGCTATACCC	AATTGAAATC	AAGGCATATC	GAGGACTACC	AAGCCTTATT	CCAGCGTGT	13140
CAATTGGATT	TGGAAGCTGA	TGTTGACGCA	TCCACTACAG	ATGATTTGTT	AAAAAATTAT	13200
AAGCCACAAG	AAGGGCAGGC	TTTGGAGGAG	CTGTTCTTCC	AGTATGGACG	GTATTTATTG	13260
ATTAGTTCGT	CCAGAGACTG	CCCAGATGCT	CTACCAGCTA	ACCTACAGGG	AGTCTGGAAT	13320
GCGGTCGACA	ATCCTCCTTG	GAATTCGGAC	TATCACTTAA	ATGTCAATCT	GCAGCTGAAT	13380
TATTGGCCAG	CCTATGTTAC	CAATCTCCTA	GAGACGGTCT	TTCCAGTCAT	CAACTATGTA	13440
GATGATTTGC	GTGTCTATGG	TCGTCTAGCG	GCTGTAAAGT	ATGCAGGAAT	CGTCTCTCAG	13500
AAAGGTGAGG	AGAATGCTTG	GTTGGTTCAT	ACTCAAGCGA	CTCCCTTTGG	TTGGACGGCA	13560
CCTGGTTGGG	ATTACTATTG	GGGTTGGTCA	CCAGCTGCCA	ATGCGTGGAT	GATGCAAACC	13620
GTTTATGAAG	CCTATTTATT	TTATAGGGAC	CAAGACTATC	TCAGGGAGAA	AATTTATCCC	13680
ATGTTGAGGG	AAACGGTTCC	TTTTTGAAT	GCCTTTTTAC	ATAAGGATCA	GCAGGCGCAG	13740
CGTTGGGTGT	CTTCTCCGTC	TTATTCCCCA	GAACATGGGC	CGATTTTCAT	TGGCAATACC	13800
TATGACCAAT	CTCTGATTTG	GCAGTTATTT	CATGATTTTA	TTCAGGCTGC	TCAGGAATTG	13860
GGACTGGATG	AGGACTTGT	GACTGAGGTT	AAGGAGAAGT	CTGATTTACT	AAATCCTTTG	13920
CAAATCACTC	AATCTGGTCG	AATCAGGGAG	TGGTATGAGG	AGGAAGAGCA	GTATTTTCAA	13980
AATGAGAAAG	TGGAGGCCCA	GCATCGGCAC	GCTTCCCATC	TAGTGGGACT	CTATCCTGGC	14040
AATCTCTTTA	GCTACAAGGG	ACAAGAGTAT	ATTGAAGCGG	CGCGTGCTAG	CCTCAATGAT	14100
CGTGAGATG	GCGGCACAGG	CTGGTCCAAG	GCTAATAAGA	TCAATCTCTG	GGCGCGTTTG	14160
GGAGATGGCA	ATCGAGCCCA	TAAATTATTG	GCAGAGCAGT	TAAAGACATC	CACCTTGCAA	14220
AATCTTTGGT	GTAGCCATCC	TCCTTTTCAG	ATAGATGGTA	ATTTTGGTGC	TACTAGTGGC	14280
ATGGCAGAAA	TGTTACTCCA	GTCTCATGCA	GCTTATCTGG	TACCTCTAGC	TGCCCTACCT	14340
GATGCTTGGT	CAACAGGTTT	TGTTTCAGGC	TTAATGGCAC	GTGGACATTT	TGAAGTGAGC	14400

367

ATGAGCTGGG AAGATAAAAA ACTCTTACAG TTGACCATT TATCAAGGAG TGGAGGAGAT	14460
TTGCGAGTTT CTTATCCAGA TATTGAGAAG AGTGTGATTA AAATGAATCA AGAAAAATA	14520
AAAGCGAAAT GCATGGGGAA AGATTGTATT TCGGTGGCAA CAGCAGAAGG TGATCTTGT	14580
CAATTTTATT TTTAAGAAGA TGTTATAAGG CAGTAATTTG AAAGTGCCTT TTAATAAGGA	14640
TTTAAGAATA TAAGCAGTTT TCAACTAGTT GAAAAACGT TATAATGATA ATAGGAAGTA	14700
ATACTCAATG AAAATCAAAG AGCACAACT AGGAAGCTAG CCGCAGGTG CTCAAAACAG	14760
TGTTTTGAGG TTGCAGATGG AAGCTGACGT GGTTTGAAGA GAGATTTTCG AGGAGTATAA	14820
TTTGTTTGAT AGAGGGTGGG TCTGATGGCT TATATTGAGA TGAAACACTG TTACAAGCGT	14880
TATCAGGTTG GGGACACGGA GATTGTGGCC AATTGTGATG TGAATTTTGA GATTGAAAAG	14940
GGGAGCTGG TTATTATCCT TGGTGCTTCA GGTGCAGGCA AGTCAACAGT TCTTAACCTT	15000
CTTGGGGGAA TGGATACCAA TGATGAAGGG GAAATCTGGA TTGATGGTGT TAATATTGCG	15060
GATTATAGTT CCCACCAGCG CACCAATTAC CGTAGAAATG ATGTGGGGTT TGTTCACAG	15120
TTTTATAATC TAGTTTCTAA TCTGACAGCT AAGGAAAATG TGGAACTGGC TTCTGAAATT	15180
GTGACAGATG CCTTGAATCC TGATCAGGCC TTGACAGATG TAGGTCTGGC TCATCGTCTC	15240
AATAACTTTC CAGCCCAGCT TTCTGGAGGG GAGCAACAGC GAGTCTCCAT TGCACGCGCG	15300
GTAGCCAAAA ATCCTAAAA TCTCCTTTGT GATGAACCGA CTGGAGCCTT GGATTATCAG	15360
ACGGGCAAGC AGGTTTGTAA AATTCTCCAA GACATGTCTC GTCAAAAGGG AGCGACGGTC	15420
ATCATCGTGA CTCATAATGG AGCTTTGGCG CCCATTGCTG ATCGCGTGAT TCAAAATGCAC	15480
GATGCCAGTG TCAAGGATGT GGTGCTCAAC CAGCATCCTC AGGATATTGA CAGTTTGGAG	15540
TACTAGCATG ATCAAGCGAA AAAGTTATTG GAAGGACTTA GTTCAGTCCT TCACAGGCTC	15600
CAAGGGGCGT TTTTATCCA TCTTGATCCT GATGATGTTG GGATCTCTAG CCTTAGTAGG	15660
CCTCAAAGTA ACCAGTCCCA ACATGGAGGC GACAGCTAAT GCTTATTTAA CAACTGCTCA	15720
AACCTTGGAT TTGGCAGTCA TGTCTAACTA TGGCTTGGAT CAAGCAGACC AAGAAGAACT	15780
AAAACAGACG GAGGGCGCAG AGGTGAGTT TGGCTATTTG ACAGATGTGA CTATGGATAA	15840
TGGGCAGGAT GCCATTCGGC TGTACTCCAA ACCAGAGCGA ATTCAACCT TTCAGCTAAG	15900
AAAGGGACGA CTTCCTCAGT CAGACAAGGA AATCGCTTTG GCCACTCATT TGCAAGGCCA	15960
ATACAGCGTG GGACAGGAGA TTAGTTTAA AGAAAAAGAA GAGGGTCATT CCTCTTTAAA	16020
AGACCATACT TATACCATTA CTGGTTTGT GGATTCGGCT GAAATCCTCT CCCAGCGAGA	16080
TATGGGCTAC GCAGGAAGTG GAAGTGGGAC TCTGACAGCC TATGGGGTGA TTTTACCTAG	16140

368

TC AATTTGAT CAGAAAGTCT ACAATATAGC TCGTTTGAAA TATCAAGATT TAGCGGGTTT	16200
AAATGCCTTT TCATCAGCTT ATGAAGAAAA ATCCAAGCAA CATCAAGAAG AGCTTGAACA	16260
AATTTTATCA GATAATGGCA AGGTACGTCT GCAACTTTTG AAAAAAGAAG GACAAGAGTC	16320
TCTAGACAAG GGGCAAGAGA CCCTTGACAA GGCTCAGACT AATTTGCAGG AAGGCAAGCG	16380
TCGTTTAGCA GCTGCTCAAG CTCGTATACA GGCTCAAGAA AGTCAACTAG CCTTGTTTCC	16440
TCAAGTTCAG AGAGAGCAGG CTAGTGCTCA ACTTACCCAA GCCAAGCAGG AATTGGGCAA	16500
GGAAGAGGAC AAACATAAGC AAGCTGAACA AAATCTAGCC CAAGAAAAGG AAAAATTAGA	16560
AAAACATCAG CAAGTCTTGG ATGATTTGGC GGAGCCAAGG TATCAGGTTT ATAATCGTCA	16620
GACCATGCCA GGTGGTCAGG GCTATCTTAT GTATAGCAAT GCTTCATCCA GTATTGAGC	16680
AGTGGGCAAT ATCTTCTCTG TGGTACTTTA TGCCGTAGCA GCCATGGTGA CCTTTACGAC	16740
CATGACTCGC TTTGTAGACG AAGAGCGAAC TCATGCAGGG ATTTTAAAGG CCTTGGGTTA	16800
TCGTAGTAAG GATATTATCG CCAAGTTTCT CCTTTATGGA CTAGTAGCTG GGAAGTCCG	16860
AACGGCTCTA GGTAGTATAC TTGGTCATTA TTTGCTAGCC AGTGTAATTT CAAGTGTCAT	16920
TACAAAAGGC ATGGTGGTGG GAGAACTCA GATTCAGTTC TATTGGACCT ATAGCTTACT	16980
AGCTTTTGTC TTGAGCTTGT TGGCGAGTGT GTTACCAGCC TATCTGGTGG CTTGGAGGGA	17040
ACTTCATGAC GAAGCAGCCC AGCTTCTACT TCCTAAACCT CCTGTCAAAG GAGCTAAAT	17100
CTTATTGGAG CGTATCGGTT TTATCTGGCG TCGTCTCAGT TTTACTCATA AGGTAACAGC	17160
CCGCAACATC TTTCGTTATA AGCAGAGAAT GTTGATGACA ATCTTTGGTG TGGCAGGTTC	17220
TGTAGCTCTG CTCTTTGCAG GTTTGGGAAT CCAATCTTCT GTAGCAGGAG TTCCGTCTAA	17280
ACAGTTTCAA CAAATCCAAC AGTATCAGAT GCTTGTCTCT GAAAATCCTA GTGCGACCAA	17340
TCAGGACAAG GTAGAGCTAG CAGAAGTGT GAAAGGGCAG GAGATACTAG CCTACCAGAA	17400
AATCTATTCT AAAGCGCTAT ACAAGGATTT CAAAGGCAAA GCTGGTCTTC AAAACATTAC	17460
TCTTATGATG ATAGAGAAGG AAGATTTGAC TCCCTTTATC CATCTTCAAC ATCATCAGCA	17520
GGAGCTGACA TTAAAAGATG GCATCGTTAT TACAGCTAAA CTCGCCCAGC TGGCAGGTGT	17580
CAAGGTTGGG CAGACTTTAG AAATTGAAGG TAAGGAACTA AAGGTCGTTG CTATTACTGA	17640
GAACTACGTT GGTCACCTTA TTTATATGAG TCAGGCTAGC TATGAGCAAC TTTACGGACA	17700
GCTACCCCAA GCCAACACTT ATCTGGTCTC ATTAAGGGAT ACCACTGCAA CTAGTATCGA	17760
AAGTCAGGCG GGCTTGCTTA TGAATCAATC TGCGGTGTCC AGCGTTGTCC AAAATGCTTC	17820
AGCCATTCTG CTCTTCGACT CTATCGCTAG CTCACTCAAT CAGACCATGA CCATCTGGT	17880
CATCGTATCG GTTCTATTAG CTATTGTCAT CCTTTACAA CTGACCAATA TCAACGTAGC	17940

TGAGAGAATC CGTGAACCTCT CCACTATCAA GGTTCCTGGT TTTCATAATA ATGAAGTCAC	18000
CCTCTACATT TACCGTGAGA CGATTGTGCT GTCCCTTGTG GGAATCGTAC TTGGTCTGAT	18060
AGCTGGTTTC TATTTACACC AATTTTGTAT TCAAATGATT TCGCCTGCCA CTATTCTCTT	18120
TTATCCGCAG GTAGGCTGGG AAGTCTATGT AATCCCAGTG GCAGCAGTAA GCATCATTTT	18180
GACCTTGCTT GGTTCCTTCG TCAATTATTA TCTGAGAAAG GTTGATATGT TAGAAGCCCT	18240
GAAATCTGTA GAGTAAGGTA GTTATTTTGA GCTGATTGAA CTTCTATTTA CTAATATTCA	18300
AAAATCCTCC GTTTCAAAGA GCAGGGAACCT CTTTGTGACA GAGGATTTTT TCTATAGGGC	18360
TTTAGCAGCT GCAATTGCGG CTTCGAAGTT TGGCTCAGAA TTGATATTAT CCACGTATTC	18420
AACGTAGCGA ATCGTATTGT CAGTATCGAG GACAAAGACT GCGCGTGCTA ATAGGTGCCA	18480
TTCGTGATC AAGAGGGCAT AATCGCGCCC GAAAGAATGG TCAAAGTAGT CTGAAAGCAT	18540
AATGGCATTG TCAAGGCCCT CAGCACCGBA CCAACGTTTT TGAGCAAAAG GTAGGTCCAT	18600
TGAAACAGTC AATACGACCG TGTGTGCCAG TCCAGCCAAT TCTTCATTAA AACGACGTGT	18660
TTGAGTTGAG CAGATGCCTG TATCGATAGA AGGAACGACA CTCAAGACTT TTTCTTGCC	18720
ATCAAAATCA GCCAGACATT TTTTAGAAAG ATCTGTTGTA GTAACAGAAA AATCAAGCGC	18780
CTTGTGCGCG ACTTGTAGTT GTTTACCTGT AAAGCTCACA GGATTTCCGA GAAAAGTTAC	18840
CATAGGATAC TCCAATCTTT TTTCTTCCAT TTTAGCTGAA ACAGTCGGAA TTTTCCAATG	18900
ATTTGACCGG AAATATGGGC ATAGAAAAAA CGCCAGCTCA TGTGAGAATG ACGTTTTTCA	18960
TAGGTTTATT TTGCCAATCC TTCAGCAATC TTGTCAAGGT TGTATTTTAT CATGCTGTAG	19020
TAGCTGTCCG CTTCTTTACC TTGTTCTGCG ATAGAGTCAG TAAAGATTG AGCGTAGATT	19080
GGGATGTTTG TGTCTTGAGA AACAGTTTTT ATTGGACGGT CATCCACACT TGATTCTACA	19140
AAGAGTGATG GAACTTTTGT TTGGCGAAGT TTTTCAACCA AGGTCTTGAT TTGTTCAAGG	19200
GTTCCTTCTT CTTCAGTATT GATTTCCCAG ATGTAAGCAC TTGGGACACC ATAGGCTTTA	19260
GAGAAGTATT TGAATGCTCC TTCGCTGGTT ACAATGAGTT TCTTTTCAGC AGGGATCTTA	19320
TTAAATTTAT CCTTACTTTC TTTATCAAGT TTGTCTAACT TATCAGTATA TTCTTTGAGA	19380
TTTTTTTCAT AGAATTCTTT ATTGTTAGGG TCTTTGGCGC TCAATTGTTT GCGGATATTT	19440
TTAGCAAAAA TAATACCGTT TTCAAGGTTA AGCCAAGCGT GTGGGTCTTC TTTTCTTTT	19500
TCATTTTGAC CTTCAAGGTA GATAACATCA ACGCCGTCGC TGAAGCGAA GTAGTCTTTG	19560
TTTTCAGTTT TCTTGCCATT TTCTACCAAT TTTGTAAACC AAGCATTGCC ACCTGTTTCA	19620
AGGTTGATAC CGTTATAGAA AATCAAATTA GCCTCAGAAG TTTTCTTAAC GTCTTCAGGA	19680



370

AGTGGTTCGT ATTCGTGTGG GTCTTGCCCA ATCGGAACGA TACTATGAAG GTCAATTTTG	19740
TCACCAGCAA TATTTTtagt AATATCAGCG ATGATTGAGT TTGTAGCAAC AACTTTTtagT	19800
TTTTGACCAG AAGTTGTATC TTTTtttCCG CTAGCACATG CTACAAGAAT GATTGCAGAA	19860
AGAAAGAGAA CGAGTAATGT ACCTAATTTT TTCATTAGAT CCTCCAATTT ATTAGGGCTT	19920
TGCCCCttAT TTTAACAAAT GTTTATTTTT CAGTTTCAAA TATCGTTGTT TGGGAGCGAT	19980
AAAGAAGCTA ATGAGAAAGA AACTAGCAGC TGTAAGCACG ATACTAGAAC CTGCCGCAAC	20040
ATTAAAACTA TAGCCAATAA AGAGTCCCAA AACTGAAGCA GTAGCTCCGA AGGTTGAGGA	20100
AAGGAAAATC ATACTTTTCA GACTATTAGC ATACAGATAA GCAGTTGCAG CTGGGGTAAAT	20160
CAGCATGGCT ACAATCAGGA TAGTTCCGAC ACTTTGCATG GCTGTCACAG ACACGAGAGT	20220
CAGGAGTACC ATGAGAAGGT AGTGATAGAA ATTGACAGGC ATTCCCATGG CTTTAGCCAA	20280
GAGTTCATCA AAGGAAGTTA TCAAGAGTTG CTTGAAGAAA ATCCAGATTA ACAAGAGGAT	20340
AGCTGCCCCC ACACCCATAG TAATAAACAT ATCCGTATCT TGGACGGCCA GGATATTACC	20400
AAAAAGGATA TGGAAAAGGT CAGTTGAACT TTTAGCGACA CCAATCAAGA TGATACCGAG	20460
GGCTAAGAAA GAAGAAAAGG TAATGCCGAT GCGGTATCG CTTTtGATAA TCGAGTTTCC	20520
TTTGATGTAG GTAATGATGA TGGCAGCTAG CAATCCAAAG ACAATGGCTC CGATAAAGAA	20580
GTCAAGGCCC AAGATGAAGG ATAGGGCTAC ACCTGGTAAG ACAGCATGTG AAATGGCATC	20640
TCCCATGAGT GACATCCCGC GTAGAATAAT GAAACATCCC ACAGCTCCAG CTACAATCCC	20700
GACGACAATA GCTGTTATCA AGGCATTTTG TAGGAAATGG AATTTTtGCA ATCCATCGAT	20760
AAATTCTGCA ATCATAGGTC ACCTCCATTG AAAAAGAGTT GATTACCGTA AGCTTCTTTT	20820
AGATTGGTTT CCGTAAAAGT TTCTTTTGGT GGACCAAAGG CAATCACTTC TCGATTGACA	20880
AGTAAGACTT GATCGAAGTA GTGGGGAATC TTGCTGAGGT CGTGGTGAAC GATGAGAACC	20940
GTCTTCCCAG CTTTTTTCAA ATCTCTCAGC GTATTtATGA TGATTTCCTC ACTGACAGAG	21000
TCAATCCCAG CAAAGGGTTC ATCCAAGAGG ATATAGTCGG CTTCTGCAC CAAACATCTG	21060
GCAATCAAGA CCCGCTGGAA TTGACCTCCA GACAGTTGAC TAATTTGACG TTCAGCGTAG	21120
TCAGCTAGGC CGACGATTTC AAGGGCCTCT TGCACTTTCT TCCAATGTTT AGCCTTTAAA	21180
CTTCGAAAGA GAGGAATAGA GGGAAATAGT CCTAACGAGA CGCATTtCCTT GACCTTGATG	21240
GGAAAGTTGT AGTCGATATT GATTTTTTGT TCGACATAGG CAATTCCGTG TAAGGATTTT	21300
TTAACTTCCT TGTCATCGAG AAATGCCTGA CCTTGATGTG GGATAATtCC CAACATACCT	21360
TTTAATAGTG TTGATTtCCC AGCGCCGTTT GGACCAATGA TGCCGGTAAT TGTTGGTCCA	21420
TGGAGCACTA GTGAAATATC CTTAAGTGCC AACGTTTCTT TGTAGGAGAC ACTGAGGTTT	21480

371

TCGATACGTA TCATAAACTT GTATTCCTCC TGTCTCTTAA TATACATTAA AAAAAAAATT	21540
AAGTCAAGTT AATTTTGGAA AAAATTAAAA TAATAACTGA AAAATAGATT CTAAGATAA	21500
CTTTCAGGAT AAATTTCTAA ATTATAAAAC GCATAGTATC AAGTGTAATA AACTTGGAAT	21660
TATGCGTTTT ATCATGGAAA GATTTTTTAT AATAGCTAAA AAATAA	21705

(2) INFORMATION FOR SEQ ID NO: 37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6171 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GATCCCCAGG AAAAACCAGG GTTTTCCCAA TCAATCGTTA CTGTCATATT CCACTCCTTA	60
TTCTAAAAAC CTATTTCTTA TATTCTACAC TATTTTCTA AAATAGCAAG TATATTTTGT	120
AATTTTCAGA AAATTTCTCC AATAAAAACC AACTCTTAGA ACTGATTCTT CATTTCACCT	180
ATTTATCTTC AGTAACTACT TCCTGAAGAT AAGCGTCAAA AACTTCTTCA TCTGAAATCG	240
TGTCAGAAAT GAAGCTTCCA TTGCTAGTGC GTTCTGACAA GTTCAAGTCT TGCAATCGGC	300
TTTCATAGAT TGTTCCTTTA TTGGATTGGA CAAGCAGAGT TTGGTCGTTT ACATCCACTT	360
CCGTACTGAA GAAATCGCCA ACAAATCCTT GCTCTGCAAC TGCTCCTGCC AAGAAGACAC	420
GATGCGGTTT GTTTTCAAC TCACGCAAGA CTTGTAATCC TCGTTTGGCA CGGCTGGTTG	480
CTAGAATTTT CTCAATGGAA ACACGTTTCA AGCTTCCACG CTGGGTCAAG AGGTAGAAGG	540
ACGAAGTATT ACAGATAAAG CCAGATTGGA GGACATCATC TTCTTTCAAA TTCATAGCCT	600
TGACACCTGC TGCCTTAGCA CCGACAACCG GAACCTCTTC GATATTGAAA CGCAGGGCAT	660
AACCATTTTG ACTAACCAAG ACAACATCAT CTAGTTTAAT CGGAGCCACT GCTACAATCT	720
GATCTGTATC GTCTTTGAGC TTAGCATACT TGACAGACTT AGATCTATAG GTCCGCCATG	780
GAGTGAATTC TTTTCGCTCT ACCCGTTTGA TTTGACCAAG GCGAGTCACT GCAAAGTAGG	840
TTGTCGCATC GTCAAACCTGA TCCAGTACTT CCACATAAAG GATTTCTTCA TTCGTTTCAA	900
AGTTTGTGAT GGTTTGGCTC AGATGCTCTC CGATGTCCTT CCAACGAATA TCTGCCAACT	960
CATGGATTGG TCTGTAGATG ACATTTCCAA GACTTGTGAA CATCAAGAGG TGCTGGGTTG	1020
TCTTGGCAGA TTGAACAAAA ATCAAACGGT CATCATCAGC CTTGCCAATT TCTTCCAAGG	1080
TGGAAGCCGC AAAGGAACGT GGAAGGTGAC GCTTGATGTA ACCTGCCCTG GTCACGCTGA	1140

372

CGTAGGTATC TTCCTCAGCG ATAAGACTAG CTGTATCAAT CTCAATTGCT TTCGCAGTGT	1200
CTTCTAAAGA ACTCAAACGA GGAGTTGCAA ATTTCTTCTT GACCTCACGA AGTTCTTTCT	1260
TCATGAGATT GTACATAGTC CTTTCATCAC CGATAATAGC CGCCAGCATA GCAATCTTCT	1320
CACGAAGCTC TGCTTCTTCT TCCTGCAAGA CAACCACATC GGTATTGGTC AAACGGTACA	1380
GTTGCAAAGT TACGATAGCC TCAGCCTGTT CTTCCGTAAA ATCATAGCTA ACTTTGAGGT	1440
TTTCCTTGGC GTCCGCCTTA TTCTCAGAAG CACGGATAAG AGCAATGACT TCATCCAAAA	1500
TCGAAATCAC ACGAATCAAA CCTTCGACGA TATGGAGACG TTTCTCAGCC TTTTCTTTGT	1560
CAAAGCGTGA ACGCGCCAAA ATCACTTCTC GACGGTGAGC GATATAGCTA GACAGGATTG	1620
GAACAATCCC AACCTGACGA GGTGTGAAAT TGTCAATCGC CACCATATTA AAGTTGTAGT	1680
TGATTTGTAG GTCGGTGAC TTAATAAGT AGTTGAGAAC AAGCTCAGTA TTAGCGTCTT	1740
TCTTAAGTTC GATAGCGATA CGAAGACCAT CACGGTCAGA CTCATCACGA ACCTCAGCAA	1800
TCCCAGCTAC CTTGTTATTA ACACGAACAT CATCGATTTT CTTGACTAGA TTGGCCTTAT	1860
TGATTTCATA AGGAATCTCA ATAATAACGA TTTGTTCTT ACCACCTTTT AGCTTTTCAA	1920
TTTCAGTCTT GGAACGAACA ACCACGCGCC CTTTCCCAGT CTCATAAGCT TTCTTGATTT	1980
CATCAGGACC CTGAATAATA GCCCCTGTAG GGAAGTCTGG TCCAGGCAAG AATTCATGA	2040
GTTTATCAAT CTTTGCAAGT GGGTGGTCAA TCATGTAAAC TGCAGCATCT ATGACCTCAG	2100
CTAAATTATG GGGAGGAATG TCTGTGGCAT AACCAGCCGA AATCCCAGTC GAACCATGA	2160
CCAAGAGGTT TGGAAAGGCT GCTGGCAAGA CCGTTGGTTC TTTCTCCGTA TCGTCAAAGT	2220
TCCATGCAAA AGGAACTGTC TTTTCTCGA TATCCTGAAG AAGGTAGCCT GCAATTTTCA	2280
ACAAACCTGC CTCAGTATAA CGCATAGCCG CAGGAGGATC TCCGTCCATA GAACCGTTAT	2340
TACCGTGCAAT TTCAACTAGA ATCTCAGAT TTTTCCAGTT CTGTGACATA CGAACCATGG	2400
CATCATAGAT AGAAGAATCC CCGTGTGGGT GGAAATTCCC CATGATGTTT CCGACTGACT	2460
TGGCCGACTT ACGGTAGCTC TTGTCAAAAG TATTGCTATC CTTATTCATA GAATAAAGAA	2520
TACGGCGCTG AACCGGCTTC AACCCATCAC GAATATCTGG CAAAGCCCGG TCTTGAATAA	2580
TGTACTTGGA GTAGCGACCA AAGCGCTCTC CCATGATGTC CTCCAGGGAC ATGTTTTGAA	2640
TGTTAGACAT AAGATACAAA GCCCATAAAA TACCAAGTGA AAATAGAAAA TTCTTGAAGT	2700
AAGCAAATCT ACAAGAGAAT TTATCTTTT CACACAGTAT CTAGGGCGTG TTCAACTCCT	2760
TTCAAAGAAT GTAGAGTAGG TTTTATGCA GTAAAAGATA TTTTACGGGA ATTCTCCCCG	2820
TGTTCAAGTA CGATAAGTAA CCAAATATC CTGTTTGTAT TTTTCAATAT GAAATCTGG	2880
TTTTCCAAAA TTAGTCTTAG TTTGTGTCTT AGCCGCTCCC TTAAGCGCCT CTTTGAGATA	2940

373

AGCACTCATA GCAGATTCTT CATTAATAAT CCTGCAATTT TTTCAAACCA AGATTTTCAA	3000
ACTGCTTTT CACATAGTCA TTCACATCCG ACTCTAATTT CCAGTTTACT AACATATTAT	3060
TTTCTTTCAT TAAAACACTG TCGTTTCTTC TAGCGTAAAC TTGACATTAT CTTCAATCCA	3120
TTTACGGCGT GGTTCACCT TATCTCCCAT GAGAACATTG ACGCGGCGTT CGGCGCGCGC	3180
TAAATCTTCA ATTGTGACAC GGATGAGGGT ACGTGTTCCT GGGTTCATGG TTGTTTCCCA	3240
GAGCTGGTCC GCATTCATCT CACCAAGTCC TTTGTATCGT TGGAGGGTAG CGCCTTTACC	3300
GAAGTGTMTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA CTTCTTCTTT	3360
CTTGCCTTTA CCTTTGGACA TCTTGTAAG AGGTGGGAGG GCAATATAGA CATGACCTGC	3420
CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT GGATATGGGC	3480
ACCGTCGGTA TCCGCATCGG TCATGATAAT GATCTTATCA TAGTTGGCAT CTTCAATAGA	3540
GAAGTCTGCT CCAACACCCG CACCAATGGT ATAAATCATG GTATTGATCT CTTCATTTTT	3600
GAGGATATCC GCCATCTTGG CCTTGGCTGT ATTGACAACC TTACCACGAA GAGGTAGAAT	3660
AGCCTGGAAC TTGCGGTCAC GACCTTGTTT GGCAGAACCA CCGGCAGAGT CCCCCTCAAC	3720
TAGATAGAGT TCATTCTTAG CAGGATTCCT AGATTGGGCT GGGGTCAATT TCCCAGACAA	3780
CAAGCCCTTA TCTTTCTTGT TTTTCTTCCC ATTTGCGCTC TCATCACGCG CTTTACGTGC	3840
TGCTTCACGA GCATCACGGG CCTTGATAGC CTGCGGATG AGGTTAGAAG CTAATTCCCC	3900
ATTTTCCATA AGGAAAAAGG TCAACTTATC AGCCACTATT CCATCCACAA CTGGGCGAGC	3960
TAGGGGGCTT CCTAGTTTAT CCTTGGTCTG TCCTTCAAAC TGCAAGTGT CTTCAGGAAC	4020
TAAGATAGAA AGAACGGCCG CTAGTCCCTC ACGATAGTCT GAACCTTCAA GGTTTTTATC	4080
TTTTTCCTTG AGAAGACCTG TTTTACGTGC ATAGTCATTC ATGACCTTGG TAATGGCAGA	4140
CTTGAGTCCT GTCTCGTGCG TTCCACCGTC CTGGTGCGA ACGTTATTGA CAAAAGATAG	4200
AATGTTATCT GAGAATCCGT CATTGTACTG GAGGGCTACT TCCACTTGAA AACCATTGTC	4260
TTCCCCTTCA AAGTAAAGAA CTGGCGTCAA GATTTCTTGA TCTTCGTTGA GATAAGAAAC	4320
AAAATCTTGT ACTCCATTCT CATAGTGGAA CTCAATCGCT TCATTTGTTT GCTTGTCCTG	4380
TAAAGACAAG GTCACATTTT TCAAGAGAAA GGCTGATTCA TTAAGGCGCT CTGAAATGGT	4440
ATTGTAATTG AAATCTGTCT TAGAAAAATAT AGTCGCGTCA GGCATAAAAG TAACTTTGGT	4500
GCCTGTTTTA GACTTGGGTG CTGTACCGAT TTTCTTCAAA GTCGTGACAG GTTTTCCACC	4560
ATTTTCGAAA CGTTGCTTGT AAAGTGGGCC ATCAGGGGTA ATTTCAACTT CTAACCAGCT	4620
AGAAAGGGCG TTAACAACGG AAGAACCAC TCCGTGAAGT CCACCTGATG TCTTATAGCC	4680

374

ACCTTGACCG AATTTCCCTC CGGCATGAAG AATCGTAAAG ATAACCTCAA CAGTTGGAAT 4740  
 TCCCATAGCG TGCATACcTG TCGGCATCCC ACGTCCATGG TCTTGAACCG TTAGACTACC 4800  
 GTCTTTATTG ATAGTTACAT CAATACGATC ACCAAACCCA GACAAGGCTT CATCGACTGC 4860  
 ATTATCAACG ATTTCCCAA CTAGGTGATG AAGACCAGCG CCATCGGTCG ATCCAATATA 4920  
 CATCCCTGGA CGTTTTCGGA CCGCATCCAA CCCTTCTAGC ACCTGAATAG CATCATCATT 4980  
 ATAATTGTTA ATATTGATTT CCTTTTTTGA CACAAGGAAC CTCCTATTCTG TTCATCTTTA 5040  
 CTATTCTACA GGTTTTCCAA GGATTTTGCA AAATTTTCT TTCTCCGATG TGACAATTTT 5100  
 AGCAGAGATT CTCTGCTTTT CTTTCCCAAT TCATGATATA ATAGGAGTAT GATTACAATA 5160  
 GTTTTATTAA TCCTAGCCTA TCTGCTGGGT TCGATTCCAT CTGGTCTCTG GATTGGACAA 5220  
 GTATTCTTTC AAATCAATCT ACGCGAGCAT GGTTCGGTA ACACTGGAAC GACCAACACC 5280  
 TTCCGCATTT TAGGTAAGAA AGCTGGTATG GCAACCTTTG TGATTGACTT TTTCAAAGGA 5340  
 ACCCTAGCAA CGCTGCTTCC GATTATTTTT CATCTACAAG GCGTTTCTCC TCTCATCTTT 5400  
 GGACTTTTGG CTGTTATCGG CCATACCTTC CCTATCTTTG CAGGATTAA AGGTGGTAAG 5460  
 GCTGTGCGAA CCAGTGCTGG AGTGATTTTC GGATTGCGC CTATCTTCTG TCTCTACCTT 5520  
 GCGATTATCT TCTTTGGAGC TCTCTATCTT GGCAGTATGA TTCACTGTC TAGTGTCA CA 5580  
 GCATCGATTG CGGCTGTTAT CGGGGTCTG CTCTTTCCAC TTTTGGTTT TATCCTGAGT 5640  
 AACTATGACT CTCTCTTCAT CGCTATTATC TTAGCACTTG CTAGTTTGAT TATCATTCGT 5700  
 CATAAGGACA ATATAGCTCG TATCAAAAAT AAACTGAAA ATTTGGTCCC TTGGGGATTG 5760  
 AACCTAACCC ATCAAGATCC TAAAAAATAA AATGCCAGTT CTGTACTGCC CCCAAACACT 5820  
 TAGACAAATA ATTTATCCAA AGGATTTAGT TCTGTACTGC ACAGGACTAA GTCCTTTTAG 5880  
 TTTTACCTTA ATTCGTTTGT TGTTGTAGTA ATCAATATAG TCTATAATGG CTTGTTCCAA 5940  
 TTGATTAAGT GATTTAAATG TTTTCTCATA GCCATAAAC ATTTGGGATT TTTAAATGCC 6000  
 AAAGAAAGAT TCCATCCTAC CGTTGTCTTG GCTGTTGCCC TTACGTGACA TGGATGCTTG 6060  
 AATTCCTTA CTCTCTAGGA ACCGATGATA AGAATCGTGT TGGTATTGCC AGCCTTGCTC 6120  
 ACTATGGAGA ATCGTATTCT CGTAGTGCTT CTCTGTGAAT GCCTGTTCCA A 6171

(2) INFORMATION FOR SEQ ID NO: 38:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18475 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

375

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

TATTACAAAT	AAAAAACCGG	AGGAGTGCTT	TATGAAAGCC	TATACTTATG	TTAAACCAGG	60
ACTTGCTTCT	TTTGTGATG	TAGACAAACC	AGTTATTCCG	AAGCCAACAG	ACGCTATTGT	120
GCGTATTGTA	AAAACCACTA	TTTGTGGAAC	AGACCTCCAT	ATTATCAAAG	GGGATGTTCC	180
TACTTGCCAA	ACTGGTACCA	TTCTTGGCCA	CGAAGGGATT	GGGATTGTTG	AAGAAGTTGG	240
CGAAGGAGTT	TCCAAC TTCA	AAAAAGGTGA	CAAGGTCTTG	ATTTCTTGCG	TCTGTGCCTG	300
TGGTAAATGC	TACTACTGTA	AAAAAGGAAT	TTATGCTCAC	TGTGAAGACG	AAGGGGGCTG	360
GATTTTCGGT	CACTTGATTG	ATGGTATGCA	GGCTGAATAT	CTACGTGTCC	CTCATGCAGA	420
TAATACTCTT	TACCATACTC	CAGAAGACTT	GTCAGATGAA	GCTTTGGTTA	TGCTGTCAGA	480
CATTCTGCCT	ACTGGATATG	AAATTGGTGT	CTTAAAAGGG	AAAGTAGAAC	CTGGTTGCAG	540
CGTAGCCATT	ATTGGTTCAG	GTCCAGTTGG	ATTGGCTGCT	CTTTTAACAG	CCCAATTCTA	600
TTCACCAGCT	AAATTGATTA	TGGTAGACCT	AGACGATAAC	CGCTTGGAAG	CTGCCCTATC	660
ATTCCGTGCG	ACTCATAAGG	TTAATTCTTC	AGACCCTGAA	AAAGCCATTA	AAGAAATTTA	720
TGATTTGACA	GATGGTCCGT	GTGTGGATGT	CGCTATCGAA	GCTGTTGGTA	TTCTTGCAAC	780
ATTTGATTTT	TGTCAAAAGA	TTATCGGTGT	AGACGGAACG	GTTGCCAACT	GTGGTGTGCA	840
TGGTAAACCA	GTTGAATTCG	ATTTAGATAA	ACTTTGGATT	CGCAACATCA	ATGTAACAAC	900
TGGTTTGGTA	TCTACAAATA	CGACTCCACA	ATTGTTGAAA	GCACTTGAAA	GTCATAAGAT	960
TGAACCGGAA	AAATTGGTAA	CTCACTATTT	CAAAC TCAGT	GAAATTGAAA	AAGCCTACGA	1020
AGTCTTCAGT	AAGGCAGCAG	ACCACCATGC	CATTAAGGTC	ATTATCGAAA	ACGATATCTC	1080
AGAAGCCTAA	GTAGTAAAAA	TATTTTTTGT	CATAAGTAAA	TAGAAATTCA	GTCATCCATC	1140
AGATGGCTGG	ATTTTTTATC	AAAAAATTAA	GAAATGAGCA	TATTTCTTTC	CTTGTCTGGC	1200
GGAAATTGGT	ATAATATACG	GTACAAAGGA	ATGAATGAAT	ATGTATCGTG	TTATAGAAAT	1260
GTACGGAGAT	TTTGAACCGT	GGTGGTTCTT	AGAAGGTTGG	GAAGAAGATA	TGTAGCAAG	1320
TAGAAAATTT	GACCAGTATT	ATGATGCTCT	CAAATACTAC	AAAAC TTGCT	GGTTTAGATT	1380
GGAAACAAGAA	TCGCCTCTTT	ATAAAAGTAG	AAGCGACTTG	ATGACCATTT	TTTGGGACCC	1440
GGAAAGACCA	CGCTGGTGTG	ATGAATGTGA	TGAGTATTTA	CAACAATACC	ATTCTTTGGC	1500
TCTTTTGCAG	GATGAGCAGG	TTATCCCAGA	CGAAAACTA	CGCTCAGGCT	ATGAAAAACA	1560
AACCAGTCAG	GAAAGGAATC	GTTCTTGCCG	TATGAAATTA	AAATAGAGAA	AAGTAACTTT	1620
TTTGAGTTG	CTTTTTTTAT	TTTTCTAACT	CTTTGCGAAT	AGTATAGGTG	AGGAGGTAAG	1680

376

TATGGTTCAA	GAAATGTCAC	AAGAAATCAT	TCGTTTCAGCT	CGGAAAAAAG	GGACGCAGGA	1740
TATCTATTTT	GTCCCTAAGT	TAGACGCCTA	TGAGCTTCAT	ATGAGGGTAG	GAGACGAGCG	1800
CTGTAAAATT	GGTAGCTATG	ATTTTGAAAA	GTTTGCAGCC	GTTATCAGTC	ACTTTAAGTT	1860
TGTGGCGGGT	ATGAATGTGG	GAGAAAAAAG	ACGTAGTCAA	CTGGGTTTCT	GTGATTATGC	1920
CTATGACCAT	AAGATAGCGT	CTCTACGTTT	ATCTACTGTA	GGCGATTATC	GGGGGCATGA	1980
GAGTTTGGTT	ATCCGTTTGT	TGCACGATGA	GGAGCAGGAC	CTGCATTTTT	GGTTTCAGGA	2040
TATTGAAGAA	TTAGGCAAGC	AGTACAGGCA	ACGGGGACTC	TATCTTTTGT	CTGGTCCGGT	2100
TGGGAGTGGT	AAGACGACCT	TGATGCATGA	ATTGTCCAAG	TCACTCTTTA	AAGACAGCA	2160
AGTTATGTCC	ATCGAAGATC	CTGTGAAAAT	CAAGCAGGAC	GACATGCTTC	AGTTGCAGTT	2220
GAACGAAGCA	ATCGGCCTAA	CCTATGAAAA	TCTAATCAAA	CTTTCCTTGC	GTCAATCGAC	2280
AGATCTCTTG	ATTATCGGAG	AAATTCGTGA	CAGCGAGACG	GCGCGTGCAG	TGGTCAGAGC	2340
TAGTTTGACA	GGTGGCAGAC	TCTTTTCAAC	CATTCACGCC	AAGAGTATCC	GAGGTGTTTA	2400
TGACCGTCTG	CTGGAGTTGG	GTGTGAGTGA	AGAAGAATTG	GCAGTTGTTT	TGCAAGGAGT	2460
CTGCTACCAG	AGATTAATCG	GGGGAGGAGG	AATCGTTGAC	TTTGCAAGCA	GAGATTATCA	2520
AGAACACCAA	GCAGCCAAGT	GGAATGAGCA	AATTGACCAG	CTTCTTAAAG	ATGGACATAT	2580
CACAAGTCTT	CAGGCTGAGA	CGGAAAAAAT	TAGCTACAGC	TAAGCAAAAA	AATATCATCA	2640
CCCTATTTAA	CAATCTCTTT	TCTAGCGGTT	TTCATCTGGT	GGAGACTATC	TCCTTTTTAG	2700
ATAGGAGTGC	TTTGTGGGAC	AAGCAGTGTG	TGACCCAGAT	GCGTGTGGGC	TTGTCTCAGG	2760
GGAAATCATT	CTCAGAAATG	ATGGAAAGTT	TGGGATGTTT	AAGTGCTATT	GTCACTCAGT	2820
TATCCCTAGC	TGAAGTTCAT	GGCAATCTCC	ACCTGAGTTT	GGGAAAGATA	GAAGAATATC	2880
TGGACAATCT	GGCTAAGGTC	AAGAAAAAAT	TGATTGAAGT	AGCGACCTAT	CCCTTGATTT	2940
TGCTGGGTTT	TCTTCTCTTA	ATTATGCTGG	GGCTACGGAA	TTACCTGCTC	CCACAACCTG	3000
ATAGTAGCAA	TATTGCCACC	CAAATTATCG	GTAATCTGCC	CCAAATTTTT	CTAGGCATGG	3060
TAGGGCTTGT	TTCCGTGCTT	GCCCTTTTAG	CACTCACTTT	TTATAAAAGA	AGTTCTAAGA	3120
TGAGTGTCTT	TTCTATCTTA	GCACGCCTTC	CCTTTATTGG	AATCTTTGTG	CAGACCTACT	3180
TGACAGCCTA	TTATGCACGT	GAATGGGGGA	ATATGATTTT	ACAGGGAATG	GAGTTGACGC	3240
AGATTTTTC	AATGATGCAG	GAACAAGGTT	CCCAGCTCTT	TAAAGAAGTC	GGTCAAGATC	3300
TGGCTCAAAC	CCTGAAAAAT	GGCCGTGAAT	TTTCTCAGAC	GATAGGAACC	TATCCTTTCT	3360
TTAGGAAGGA	ATTGAGTCTC	ATCATAGAGT	ATGGGGAAGT	TAAGTCCAAG	CTGGGTAGTG	3420
AGTTGGAAAT	CTATGCTGAA	AAAACCTGGG	AAGCCTTTTT	TACCCGAGTC	AACCGCACCA	3480

377

TGAATTTGGT GCAGCCACTG GTTTTATCT TTGTGGCACT GATTATCGTT TTA	3540
CGGCAATGCT CATGCCCATG TATCAAATA TGGAGGTAAA TTTTAAAT GAAAAAATG	3600
ATGACATTCT TGA AAAAGC TAAGGTAAA GCTTTTACAT TGGTGGAGAT GTTGGTGGTC	3660
TTGCTGATTA TCAGCGTGCT TTTCTTGCTC TTTGTACCTA ATCTGACCAA GCAAAAAGAA	3720
GCAGTCAATG ACAAAGGAAA AGCAGCTGTT GTTAAGGTGG TGGAAAGCCA GGCAGA	3780
TATAGCTTAG AAAAGAATGA AGATGCTAGC CTAAGAAAGT TACAAGCAGA TGGACGCATC	3840
ACGGAAGAAC AGGCTAAAGC TTATAAGAA TACAATGATA AAAATGGAGG AGCAATCGT	3900
AAAGTCAATG ATTAAGGCCT TTACCATGCT GGAAAGTCTC TTGGTTTTGG GACTTGTGAG	3960
TATCCTTGCC TTGGGCTTGT CCGGCTCTGT CCAGTCCACT TTTTCAGCGG TAGAGGAACA	4020
GATTTTCTTT ATGGAGTTTG AAGAACTCTA TCGGGAAACC CAAAAACGCA GTGTAGCCAG	4080
TCAGCAAAAG ACTAGTCTGA ACTTAGATGG GCAGACGCTT AGCAATGGCA GTCAAAAGTT	4140
GCCAGTCCCT AAAGGAATTC AGGCCCCATC AGGCCAAAGT ATTACATTTG ACCGAGCTGG	4200
GGGCAATTCG TCCCTGGCTA AGGTTGAATT TCAGACCAGT AAAGGAGCGA TTCGCTATCA	4260
ATTATATCTA GGAAATGGAA AAATTAACG CATTAAAGGAA AAAAAAATT AGGGCAGTGA	4320
TTTTACTGGA AGCAGTAGTC GCTCTAGCTA TCTTTGCCAG CATTGCCACC CTCCTTTTGG	4380
GACAAATCA AAAAAATAGG CAAGAGGAAG CAAAAATCTT GCAAAAGGAA GAAGTCTTGA	4440
GGGTAGCTAA GATGGCCCTG CAGACGGGGC AAAATCAGGT AAGCATCAAC GGAGTTGAGA	4500
TTCAGGTATT TTCTAGTGAA AAAGGATTGG AGGTCTACCA TGGTTCAGAA CAGTTGTTGG	4560
CAATCAAAGA GCCATAAGGT CAAGGCTTTT ACCTTGTTAG AATCCCTGCT TGCCCTCATT	4620
GTCATCAGTG GGGGATTACT CCTTTTCAA GCTATGAGTC AGCTCCTCAT TTCAGAA	4680
CGCTACCAGC AACAAAGCGA GCAAAAGGAG TGGCTCTTGT TTGTGGACCA ACTTGAGGTA	4740
GAATTAGACC GTTCGCAGTT CGAAAAAGTA GAAGGCAATC GCCTATACAT GAAGCAAGAT	4800
GGCAAGGACA TCGCCATCGG TAAGTCAAAG TCAGATGATT TCCGTAAAAC GAATGCTCGT	4860
GGTCGAGGTT ATCAGCCTAT GGTTTATGGA CTCAAATCTG TACGGATTAC AGAGGACAAT	4920
CAACTGGTTC GCTTTCATTT CCAGTTCCAA AAAGGCTTAG AAAGGGAGTT CATCTATCGT	4980
GTGGAAAAAG AAAAAAGTTA AGGCAGGTGT TCTCCTCTAC GCAGTCACCA TAGCAGCCAT	5040
CTTTAGTCTT TTGTTGCAAT TTTATTTGAA CCGACAAGTC GCCCACTATC AAGACTATGC	5100
TTTGAATAAA GAAAAATTGG TTGCTTTTGC TATGGCTAAA CGAACCAGG ATAAGGTTGA	5160
GCAAGAAAGT GGGGAACACT TTTTAAATCT AGGTCAGGTA AGCTATCAAA ACAAGAAAC	5220



378

TGGCTTAGTG ACGAGGGTTC GTACGGATAA GAGCCAATAT GAGTTTCTGT TTCCTTCAGT	5280
CAAAATCAAA GAAGAGAAAA GAGATAAAAA GGAAGAGGTA GCGACCGATT CAAGCGAAAA	5340
AGTGGAGAAG AAAAAATCAG AAGAGAAGCC TGAAGAAGAA GAGAATTCAT AGTCAATTCA	5400
ACTATAATGC GTTGAATCCA GAATAGTCCA CTGTAGTTTC TAGAAAATTG CTGGAAATGG	5460
ATGTTAAGCT CCAATTCATT TGTTTATATC TTATTTTCAGT TTACTATACT TTGTGCTAAA	5520
TTAAAGATAT GAAACATGAT TTTAACCACA AAGCAGAAAC TTTCGATTCC CCTAAAAATA	5580
TCTTCCTCGC AAACCTGGTA TGTCAAGCAG CCGAGAAACA GATTGATCTT CTATCAGACA	5640
AAGAAATTTT AGATTTCGGT GGTGGCACGG GTCTATTAGC CTGCCCCCTA ACCCCTAGCC	5700
AAGCAGGCTA AGTCAGTCAC TCTTGTAGAC ATTTCTGAGA AAATGTTGGA GCAAGCTCGT	5760
TTGAAAGTGG AGCAGCAAGC AATCAAGAAT ATCCAGTTTT TGGAGCAAGA TTTACCGAAA	5820
AATCCCTTGG AGAAAGAGTT TGATTGCCTT GCTGTTAGTC GGGTTCTTCA TCATATGCCT	5880
GATTTGGATG CGGCTCTCTC ACTGTTTCAT CAACATTTGA AGGAAGATGG GAAACTCATC	5940
ATTGCTGATT TTACCAAGAC AGAAGCTAAT CATCATGGAT TTGATTTAGC TGAAGTGGAA	6000
AACAAGCTAA TTGAGCATGG TTTTTCATCT GTGCATAGTC AGATTCTCTA TAGTGCTGAA	6060
GACCTGTTTC AAGGAAATCA CTCAGAATTC TTTTAAATAG TAGCCCAAAA ATCACTCGCC	6120
TAGTCAGGGA GTGATTTTTT TATAAGGATG GAAAAAAGAA GGGAAATTTG GTAAGATAGG	6180
AATATGGATT TTGAAAAAT TGAACAAGCT TATACCTATT TACTAGAGAA TGTCCAAGTC	6240
ATCCAAAGTG ATTTGGCGAC CAACTTTTAT GACGCCTTGG TGGAGCAAAA TAGCATCTAT	6300
CTGGATGGTG AAACCTGAGCT AAACCAGGTC AAGGAGAACA ATCAAACCCCT TAAGCGTTTA	6360
GCACTACGCA AAGAAGAATG GCTCAAGACC TACCAGTTTC TCTTGATGAA GGCTGGGCAA	6420
ACAGAACCCCT TGCAGGCCAA TCACCAGTTT ACACCGGATG CTATTGCTTT GCTTTTGGTG	6480
TTTATTGTGG AACAGTTGTT TAAAGAGGAG GAAATTACTA TCCTCGAAAT GGGTTCTGGG	6540
ATGGGAATTC TAGGCGCTAT TTTCTTGACC TCGCTTACTA AAAAGGTGGA TTAATTGGGA	6600
ATGGAAGTGG ATGATTTGCT GATTGATCTG GCAGCTAGCA TGGCAGATGT AATTGGTTTG	6660
CAGGCTGGCT TTGTCCAAGG AGATGCCGTT CCCCCACAAA TGCTCAAAGA AAGCGATGTG	6720
GTCATCAGTG ACTTGCCTGT CGGCTATTAT CCTGATGATG CCGTTGCGTC GCGCCATCAA	6780
GTGCTTCTA GCCAAGAACA TACTTACGCC CATCACTTGC TCATGGAACA AGGGCTTAAG	6840
TACCTCAAGT CAGACGGATA CGCTATTTTT CTAGCTCCGA GTGATTTGTT GACCAGTCCT	6900
CAAAGTGATT TGTAAAAGA ATGGCTGAAA GAAGAGGCGA GTCTGGTTGC TATGATTAGT	6960
CTGCCTGAAA ATCTCTTTGC TAATGCCAAA CAATCTAAGA CTATTTTAT CTACAGAAG	7020

379

AAAAATGAAA TAGCAGTAGA GCCTTTTGT TATCCACTTG CTAGCTTGCA AGATGCAAGT	7080
GTTTTAATGA AATTTAAAGA AAATTTTCAA AAATGGACTC AAGGTACTGA AATATAAAAT	7140
AGATTTTGTT ATAATAGTTG AAAACGCTTA AAAAGGGGTA TCATGTTATG AAAAAACAA	7200
TTGCAATCAA TGCAGGAAGT TCAAGTTTGA AATGGCAATT ATACTTAATG CCAGAAGAAA	7260
AAGTATTGGC GAAAGGTTTG ATTGAACGTA TCGGTTTGAA AGATTCAATT TCAACTGTAA	7320
AATTTGACGG CCGTTCTGAA CAACAAATTT TGGATATTGA AAATCATATA CAAGCCGTTA	7380
AAATTTTATT GGATGACTTG ATTCGTTTCG ATATTATCAA GGCTTATGAC GAGATTACAG	7440
GTGTTGGACA TCGTGTGTG GCTGGTGGAG AATATTTCAA AGAATCAACA GTTGTGAGG	7500
GAGATGTTTT AGAAAAAGTT GAAGAGTTGA GTTGTGTTGGC TCCTCTACAC AACCCGGCCA	7560
ATGCAGCAGG TGTTTCGTGCC TTCAAGGAAT TGTGTCGAGA CATTACCAGT GTAGTTGTTT	7620
TTGATACTTC CTTCCACACA AGTATGCCAG AGAAAGCTTA TCGCTACCCT CTACCAACAA	7680
AATATTACAC AGAAAAACAAG GTTCGTAAAT ACGGTGCTCA TGGTACAAGT CACCAGTTTG	7740
TAGCAGGAGA AGCTGCAAAA CTCTTGGGAC GTCCATTAGA AGACTTGAAG TTAATTACCT	7800
GTCATATTGG TAACGGAGGC TCAATTACAG CTGTGAAAGC CGGCAAATCT GTAGACACTT	7860
CTATGGGGTT CACTCCTCTT GGTGGTATTA TGATGGGAAC GCGTACAGGG GATATTGATC	7920
CAGCTATCAT TCCTTATTTA ATGCAATATA CAGAGGATTT TAACACACCA GAAGATATCA	7980
GTCTGTCTCT TAACCGTGAA TCAGGTCTTT TGGGAGTTTC TGCTAATTCT AGCGATATGC	8040
GCGATATAGA AGCAGCTGTA GCAGAAGGGA ATCAGCAGGC TAGCTTGGCT TATGAAATGT	8100
ATGTTGACCG TATCCAAAAA CATATCGGTC AGTACCTTGC AGTGCTAAAT GGAGCAGATG	8160
CCATTGTTTT CACAGCAGGT GTCGGTGAAA ATGCAGAGAG TTTCCGTCGT GATGTAATCT	8220
CAGGGATTTC GTGGTTTGGT TGTGATGTTG ATGATGAAAA GAATGTCTTT GCGGTACAG	8280
GAGACATCTC AACAGAGGCA GCTAAAAATCC GTGTCTTGGT TATTCCAACA GATGAAGAAT	8340
TAGTCATTGC CCGTGACGTT GAACGCTTGA AAAAAATAAGT GAAACTAAAA AAATATTCAA	8400
TACAAGGAGT TGGGAAAGTT ATTTTCCAG CTCTTTTTTC TGATGAAATT GTCCAAAACC	8460
TTGCTATGAT TGGCTTTTTT GAAAAATATG GTATAATAGT AGTAATTTAA TAGATGGAGT	8520
TGAGTTTGA AGAAAAACTT TCGTGAAAAA AGAGAGAAAG ATTTTAAGGC GATTTTCAAG	8580
GAGGGGACAA GTTTTGCTAA TCGCAAATTT GTGGTCTACC AATTAGAAAA CCAGAAAAAC	8640
CGTTTTCGAG TAGGTCTATC AGTTAGCAAA AAAGTGGGGA ATGCCGTCAC TAGAAATCAA	8700
ATTAAGCGAC GGATTGCGCA TATTATCCAG AATGCAAAAG GGAGTCTGGT AGAAGATGTC	8760

380

GACTTTGTTG TCATTGCTCG AAAAGGAGTC GAAACCTTGG GATAGC TAGA GATGGAGAAA	8920
AATCTACTCC ATGTATTAAA ATTATCAAAG ATTTACCGGG AAGGAAATGG GAGTGAAAAA	8880
GAAACTAAAG TTGACTAGTT TGCTAGGACT GTCTCTGTTA ATCATGACAG CCTGTGCGAC	8940
TAATGGGGTA ACTAGCGATA TTACAGCCGA ATCGGCTGAT TTTTGGAGTA AATTGGTTTA	9000
CTTCTTTGCG GAAATCATTG GCTTTTATC GTTTGATATT AGTATCGGAG TGGGGATTAT	9060
TCTCTTTACG GTCTTGATTG GTACAGTCCT CTTGCCAGTC TTTCAGGTGC AAATGGTGGC	9120
TTCTAGGAAA ATGCAGGAAG CTCAGCCACG CATTAAAGGCG CTTCGAGAAC AATATCCAGG	9180
TCGAGATATG GAAAGCAGAA CCAAAGTAGA GCAGGAAATG CGTAAAGTAT TTAAGAAAT	9240
GGGTGTCAGA CAGTCAGACT CTCTTTGGCC GATTTTGATT CAGATGCCGG TTATTTTGGC	9300
CCTGTTCCAA GCCCTATCAA GAGTTGACTT TTAAAGACA GGTCATTTCT TATGGATTAA	9360
CCTTGGTAGT GTGGATACAA CCCTTGTCT TCCGATTTTA GCAGCAGTAT TCACCTTTTT	9420
AAGTACTTGG TTGTCCAACA AAGCTTTGTC TGAGCGAAAT GGCCTACCA CTGCGATGAT	9480
GTATGGGATT CCAGTCTTGA TTTTATCTT TGCAGTTTAT GCSCCAGGTG GAGTCGCCCT	9540
ATACTGGACA GTGTCTAATG CTTATCAAGT CTTGCAAACC TATTTCTTGA ATAATCCATT	9600
CAAGATTATC GCAGAGCGCG AGGCCGTAGT ACAGGCACAA AAACATTTGG AAAATAGAAA	9660
AAGAAAAGCC AAGAAAAAGG CTCAGAAAAC GAAATAAATA AGGAGGAATC TGGTAGTGGT	9720
AGTATTTACA GGTCAACTG TTGAAGAAGC AATCCAGAAA GGATTGAAAG AATTAGATAT	9780
TCCAAGAATG AAGGCTCATA TCAAAGTCAT TTCTAGGGAG AAAAAAGGCT TTCTTGGTCT	9840
ATTTGGTAAA AAACCAGCCC AAGTGGATAT TGAAGCGATT AGTGAAACGA CTGTTGTCAA	9900
AGCAAATCAA CAGGTAGTAA AAGGCGTTCC GAAAAAATC AATGATTTGA ACGAGCCTGT	9960
GAAGACGGTT AGTGAAGAAA CCGTTGACCT TGGTCATGTG GTTGATGCTA TTAaaaaaat	10020
AGAGGAAGAA GGTCAAGGTA TTTCTGATGA AGTCAAGGCT GAAATCTTAA AACATGAAAG	10080
ACATGCCAGC ACTATCTTAG AAGAAACTGG TCACATTGAG ATTTTAAATG AACTTCAAAI	10140
CGAGGAAGCG ATGAGGGAAG AAGCAGGCGC TGATGACCTT GAAACTGAGC AAGACCAAGC	10200
TGAAAGTCAA GAACTAGAAG ACTTGGGCTT GAAAGTTGAA ACGAACTTTG ATATTGAACA	10260
AGTAGCTACG GAAGTAATGG CTTATGTTCA AACGATTATT GATGACATGG ATGTTGAGGC	10320
TACACTTTCA AATGATTATA ACCGTCGTAG CATCAATCTA CAAATTGACA CCAACGAACC	10380
AGGTCGTATT ATCGGCTACC ATGGTAAAGT CTTGAAGGCC TTGCAACTGT TGGCTCAAAA	10440
TTATCTTTAC AACCGCTATT CCAGAACCCT CTACGTTACA ATCAATGTCA ATGATTATGT	10500
CGAACACCST GCAGAAGTCT TGCAGACCTA TGCGCAAAAA TTGGCGACTC GTGTTTGGGA	10560

381

AGAAGGGCGC AGTCATAAAA CAGATCCAAT GTCAAATAGC GAACGCAAGA TTATCCATCG	10620
TATTATTTCA CGTATGGATG GCGTGACTAG TTA CTCTGAA GGTGATGAGC CAAATCGCTA	10680
TGTTGTTGTA GATACAGAAT AAGTAAATC AGGTTTATCC TGATTTTGTG CTAGTTAGAG	10740
GAGGTTAAAC TGATGTTGAA TAAGATAAGA GACTATTTAG ACTTTGCTGG TTTGCAGTAC	10800
CGTAATCCTG ATAAAGCGGG AGCAGAGCGA GAGAAGATGC TGGCATTCCG CCACAAAGGA	10860
CAAGAGGCCC GAAAGGTTTT TACAGAACTG GCCAAAGCCT TTCAAGCAAG CCATCCAGAA	10920
TGGCAACTCC AACAGACTAG CCAGTGGATG AATCAGGCCC AGCGTTTGAG ACCACATTTT	10980
TGGGTTTATC TACAGAGAGA CGGACAAGTG ACAGAACCTA TGATGGCCTT ACGTTTGTAT	11040
GGGACATCTA CTGACTTTGG AATTCTTTG GAAGTCAGTT TCATCGAACG TAAGAAGGAT	11100
GAGCAAACAC TGGGCAAGCA GGCCAAAGTT TTAGACATTC CAACCGTTAA AGGGATTTAT	11160
TATCTAACCT ACTCTAATGG TCAAAGTCAA CGGTGGGAGG CGAATGAAGA AAAGCGTCGT	11220
ACTTTACGCG AGAAGGTGAG AAGTCAAGAA GTTCGAAAAG TTTTAGTGAA GGTAGATGTT	11280
CCTATGACAG AAAATTCGTC TGAAGAAGAA ATCGTAGAAG GCTTATTGAA GTCTTATTCT	11340
AAAATCTTTC CCTATTATCT AGCTACGAGA AAATAAGATA ATTTGTAAAA CATCATAAAT	11400
CATACAGTCC AAGAGTGAAC AGTCCGCTGT GTAATCTTTC GTCTTTTGTG TTGCGCTTTC	11460
GCATTATATA ATAAACTTAC AAAAACAATT CAAAAGGAGA ACAATTATGG AAGTCGTTTC	11520
AAGTGTTCTA AATTGGTTTT CTAGCAATAT TTTGCAGAAT CCCGCATTTT TCGTAGGTTT	11580
ATTGGTGTTG ATAGGATATG CACTTTTGAA AAAACCTGCC CATGACGTTT TTTCAGGGTT	11640
TGTTAAAGCA ACACTAGGGT ATATGTTGCT TAACGTGGGT GCTGGTGGTT TGGTTACAAC	11700
CTTTCGTCCA ATCTTAGCAG CTCTTAACTA CAAATTCCAA ATTGGTGCAG CGGTTATCGA	11760
CCCTTACTTT GGACTTGCTG CAGCAAACAA CAAAATTGTA GCAGAGTTTC CAGATTTTGT	11820
TGGAAGTGA ACTACAGCTC TATTGATTGG TTTTGGGAATA AATATCTTGC TCGTAGCTCT	11880
TCGAAAGATT ACGAAGGTAA GAACCTCTT TATTACTGGT CACATCATGG TACAACAAGC	11940
TGCAACAGTA TCTCTTATGG TTCTATTCTT AGTACCACAA TTGCGCAATG CTTACGGTAC	12000
AGCAGCGATT GGTATCATCT GTGGACTTTA CTGGGCAGTT AGTTCAAATA TGACTGTTGA	12060
GGCAACTCAA CGCTTGACTG GTGGTGGCGG ATTTGCGATT GGTCAACCAAC AGCAATTTGC	12120
AATCTGGTTT GTAGATAAAG TAGCAGGACG CTTTGGTAAG AAAGAAGAAA GTTTAGACAA	12180
TCTTAAATTA CCTAAGTCC TCTCAATCTT CCACGATACA GTTGTGTCAT CTGCTACCTT	12240
GATGCTCGTA TTCTTCGGAG CCATTCTTTT AATCTTGGGT CCAGACATTA TGTCTAATAA	12300

382

AGAAGTCATC ACTTCAGGAA CTCTATTCAA TCCTGCTAAA CAAGATTTCT TTATGTACAT	12360
TATCCAAACA GCCTTTACCT TCTCAGTTTA CTTGTTGCTT TTGATGCAAG GTGTCCGAAT	12420
GTTCTATCT GAGTTGACAA ACGCCTTCCA AGGTATTTC AACAATTGT TGCCAGGTTT	12480
ATTCCCAGCG GTTGACGTTG CAGCTTCTTA TGGATTGGT TCTCCAAATG CTGTCTTGTC	12540
AGGATTTACC TTTGGTTTGA TTGGTCAATT GATTACAATT GTTTTGCTCA TCGTCTTTAA	12600
AAATCCGATT CTTATTATTA CAGGATTGT ACCAGTGTTC TTGACAATG CAGCCATTGC	12660
GGTCTACGCT GATAACGCG GCGGATGGAA AGCGGCTGTT ATCCTTTCTT TTATATCAGG	12720
TGTCCTTCAA GTTGCTCTAG GAGCTCTTTG TGTGGCCCTT CTCGATTGG CATCTTATGG	12780
TGGCTACCAT GGAAATATCG ACTTTGAATT CCCATGGCTT GGATTGGAT ATATCTTCAA	12840
ATACCTTGGT ATGTTGGTT ATGTACTGT GTGTCTCTTC TTGCTTGTTA TTCCTCAACT	12900
TCAATTTGCC AAAGCAAAG ATAAAGAGAA ATATTACAAC GGTGAAGTTC AAGAAGAAGC	12960
TTAGTATCTA GAAAAGGAGA AATAAAATGG TTAAAGTATT AGCAGCGTGC GGAAATGGAA	13020
TGGGTTTATC AATGGTTATC AAGATGAAGG TTGAAAATGC TCTCCGTAAG CTTAATCAAA	13080
CAGATTTTAC AGTCAATTCA TGCAGTGTG GTGAAGCTAA AGGTTTAGCA GTAGGATATG	13140
ACATCGTAAT CGCTTCTCTT CATTTGATTC AAGAATTGGA AGGGCGAACT AATGGGAAGT	13200
TAATTTGGCT TGATAACTTG ATGGATGATA AAGAAATCAC CGAAAACTC AGTCAAGCAC	13260
TACAGTAAAA GGTGAGGAGG GGCTGGACAG AAAGTGAGAG TTATCGTTTC TGTCTTCTC	13320
CCTCTTTAAA TAAAGGAGGC AGATATGAAT TTAAACAAG CTTTAATTGA CAATGACTCG	13380
ATCCGACTAG GTTTAGAGGC TAACAATTGG AAAGAAGCAC TCAAGGTAGC AGTAGATCCC	13440
TTAATTGAAA GTGGGGCAAT TTTGCCAGAG TATTACGATG CTATCATTGA ATCGACTGAA	13500
GAGTATGGGC CTTACTATAT CTTGATGCCA GGTATGGCTA TGCCCCACGC TAGACCTGAA	13560
GCAGGTGTGC AAAGTGATGC CTTTTCATTG ATTACCTTAC AAAATCCTGT TGTATTTTCA	13620
GATGGGAAAG AGGTATCTGT TTTGTTGGCA CTAGCAGCAA CAAGTTCAAA AATTCACACA	13680
AGTGTAGCCA TTCCACAAAT TATTGCCCTA TTTGAATTAG AAGATTCTAT TGCACGTTTA	13740
CAGGCTTGCC AGACTAAAGA AGATGTCTTG GCTATGATTG AAGAACTAA GGATAGCCCT	13800
TATCTCGAAG GATTGGATTG GGAAAGTTAG AAAGAGGAAT AAAGAAATGA CAAAAAGAAT	13860
ACCTAATTTA CAAGTTGCAT TAGACCATTC AGACTTGCAA GGAGCGATTA AAGCAGCTGT	13920
TTCTCTTGGT CAGGAAGTAG ATATTATCGA AGCTGGAAT GTTTGCTTGC TTCAAGTTGG	13980
AAGTGAAGT GCTGAAGTCT TGCCTAGCCT TTTCCAGAT AAGATTATTG TGGCAGACAC	14040
AAAATGTGCT GATGCTGGTG GAACAGTTGC TAAAAATAAT GCGGTTCTGT GAGCAGACTG	14100

383

GATGACTTGT ATCTGTTGTG CAACCATCCC TACTATGGAA GCAGCTCTAA AGGCTATCAA	14160
GA CTGAACGA GGAGAACGAG GCGAAATCCA GATCGAGCTT TATGGCGATT GGACTTTTGA	14220
ACAAGCTCAG CTTTGGCTAG ATGCAGGTAT CTCACAAGCT ATTTATCACC AATCTCGTGA	14280
TGCTCTTCTT GCTGGTGAAA CTTGGGGTGA AAAAGACCTT AATAAGGTTA AAAAACTCAT	14340
TGACATGGGC TTCCGTGTAT CTGTAACAGG TGGTCTAGAT GTAGATACTC TCAAACCTCTT	14400
TGAAGGTATT GATGTCTTTA CCTTTATCGC AGGTCGTGGA ATTACAGAGG CTGTGGATCC	14460
AGCAGGAGCA GCGCGTGCCCT TCAAGGATGA AATCAAACGA ATTTGGGGGT AAATCATGGT	14520
ACGTCCAATT GGAATTTATG AAAAGGCAAC CCCAACACAC TGTACTTGCC TAGAACGTTT	14580
AAATTTTGCC AAGGAGTTAG GCTTTGATTT TGTGAGATG TCTATTGACG AACGTGACGA	14640
GCGTTTAGCA AGACTTGACT GGAGTAAGGA AGAACGCTTG GAAGTTGTCA AAGCAATCTA	14700
TGAACTGGT GTTCGTATTC CTTCTATCTG TTTTTCAGGC CATCGTCGCT ACCCATTTGGG	14760
TTCAAAAGAT CCAGTTCTAG AGGAAAAATC TCTAGAACTC ATGAAAAAAT GTATCGAATT	14820
AGCTCAAGAC TTGGGAGTTC GTACGATTCA ATTAGCTGGT TACGATGTTT ACTATGAGGA	14880
AAAGTCACCC CAGACACGCC AACGTTTAT CAAAAATTG AGAAAAGCCT GTGACTGGGC	14940
TGAAGAAGCT CAGGTGGTAC TTGCTATTGA AATTATGGAT GATCCTTTCA TCAGTAGCAT	15000
CGAAAAATAT TTGGCTATAG AAAAAGAGAT TGA CTCTCCC TTCCTCTTTG TATATCCAGA	15060
TATTGGTAAT GTGTCTGCAT GGCATAATGA TATCTATAGT GAGTTTTATC TTGGTCATCA	15120
TGCCATCGCA GCTCTCCATC TCAAGGATAC TTATGCAGTG ACAGAAAGTT CAAAGGGCCA	15180
GTTCGGAGAT GTACCTTTCG GGCAAGGTTG TGTCAAATGG GAAGAAGCTT TCGATATTTT	15240
AAAGGAAACC AATTATAATG GACCTTTCCT AATCGAAATG TGGTCTGAAA ATTGTGAAAC	15300
AGTAGAAGAA ACACGCGCAG CCATTCAAGA GCGCAAGCT TTTCTCTATC CACTCATTAA	15360
GAAAGCAGGT TTGATGTAAG ATGAATCAAG TAATCAATGC TATGCGTAAA CGACTCTGTG	15420
ATGCCAATCA ATCATTGCCA AAACATGGAC TTGTCAAATT TACCTGGGGG AATGTATCTG	15480
AAGTTAATCG CGAACTCGGT GTCATTGTTA TCAAACCATC AGGCGTGGAT TATGACGAAT	15540
TGACACCTGA AAACATGGTA GTGACTGATC TAGATGGTAA GATCCTAGAA GGGGATTTAA	15600
GACCATCTTC CGACCTCCCA ACTCATGTGC AATTATATAA GACTTGGTCA GAAATTGGTA	15660
GTGTGGTTCA CACCCATTCTG ACAGAAGCTG TTGGTTGGGC TCAGGCAGGT CGTGATATTC	15720
CTTTCTACGG AACAACCCAT GCAGATTATT TCTACGGTTC AATCCCTTGC GCCCGTAGTT	15780
TGACCAAGGA CGAAGTAGAA GTGGCCTATG AAAAAAGATC TGGCCTGGTT ATCGTAGAAG	15840

384

AGTTTGAACA TCGCGGACTT AACCCGGTTG AAGTACCAGG AATTGTTGTA CGCAATCACG	15900
GTCCATTTCAC CTGGGGCAAA AATCCAGAGA ATGCTGTTTA TCACTCTGTC GTACTAGAGG	15960
AAGTATCAAA GATGAATCGC TTTACAGAAC AAATCAATCC AAGAGTTGGA CCTGCTCCCC	16020
AGTACATACT AGAAAAACAC TACCAACGTA AACATGGACC AAATGCTTAT TATGGTCAAA	16080
AGTAAGAACG ATGAAGGAGG AGAAAAAGAT AAATTTAGCT CCTCTTTTTA CATTTGATTT	16140
TTATTGAGAG TAAAGTTGGA GTTGAAGTAA TTTTAAAGA TTTTITAGAA ATAGCGCTTG	16200
ATATATATAT GGTAAAATAA AAAGAATTGC TGTGATATCA ATAGATTGGG GGGATTTTTT	16260
AATATGGTAC TGGATAAGGC AAGTTGTGAT TTGCTTCAAT ATTTGATGGA TCAAGAAACG	16320
TCCAAAACGA TTATGGCGAT TTCGAAAGAT TTGAAAGAGT CAAGAAGGAA AATTTATTAT	16380
CACATTGACA AAATCAATGC TGCTCTGGGT GACGAGGCGC TTCACATCAT TAGTATTCCA	16440
CGAATTGGTA TTCACTTAAC GGAAGAGCAG AGAGATGCTT GTTGTAAGT ATTATCGGAA	16500
GTAGATTCTG ACGATTATAT CATGAGTGGC CATGAACGTA TGATGATAAT GTTACTATGG	16560
ATAGGTATTT CTAAAGAACG TATTACGATT GAAAAATTGA TAGAGTTAAC AGAGGTATCT	16620
AGGAATACTG TTCTCAATGA TTTGAATAGT ATTCGTTATC AACTAACTTT GGAACAATAT	16680
CAGGTGATCT TGCAAGTGAG CAAGTCACAG GGATACAACC TTCATGCCCA CCTCTTAAT	16740
AAAATTCAGT ATCTTCAATC GCTTCTATAT CATATTTTTA TGAAGAAAA TGCCACTTTT	16800
GTATCTATTT TAGAAGATAA GATGAAAGAG AGGTTAGATG ATGAGTGTTT GCTTCTGTT	16860
GAAATGAACC AATTTTTTAA GGAACAGGTT CCTTTAGTTG AACAAGATTT AGGGAAGAAA	16920
ATAAACCATC ATGAAATAAC TTTTATGTTG CAGGTTCTAC CTTATTTGCT GTTAAGCTGT	16980
CATAATGTTG AACAGTATCA AGAAAGACAT CAGGATATAG AGAAAGAATT TTCTTTGATA	17040
AGAAAAAGAA TAGAGTATCA GGTGTCTAAG AAATTAGGAG AACGTTGTT TCAAAAGTTT	17100
GAAATTTCTT TGTCAGGACT TGAAGTTTCT CTTGTAGCTG TTCTCCTCCT CTCCTATCGT	17160
AAAGATTGGG ATATTCATGC AGAAAGTGAT GATTTTCGGC AATTAAAACT TGCTTTAGAA	17220
GAATTTATCT GGTATTTTGA ATCACAAATC CGAATGGAGA TTGAGAACAA GGATGATTTG	17280
TTACGAAATT TGATGATCCA CTGTAAAGCC TTGTTATTTA GAAAGACTTA CGGTATTTTT	17340
TCTAAAAATC CTCTAACAAA ACAAATTEGA TCCAAGTATG GAGAATTATT TTTAGTCACT	17400
AGAAAATCTG CGGAAATTTT AGAAGGAGCA TGGTTTATTC GGCTAACAGA CGATGATATT	17460
GCCTATTTGA CGATTCATAT TGGAGGATTT TTAATAATA CACCATCATC TCAAAAAAAT	17520
ATGAAAAAAG TTTATCTCGT TTGTGATGAA GGTGTTGCGG TTTGAGACT TTTGCTGAAA	17580
CAATGCAAAC TTTATTTTCC AAATGAGCAA ATTGACACTG TATTTACAAC AGAACAATTT	17640

385

AAGAGTGTGG AAGATATTGC ACAAGTTGAT GTAGTGATTA CTACTAATGA TGATTGTGGAT	17700
AGCAGATTTTC CGATTTTAAAG GGTTAATCCT ATCCTTGAAG CAGAAGATAT TTTGAAAATG	17760
CTAGACTATC TTAAACACAA TATATTTCGT AATAAGAGCA AAAGTTTCAG TGAAAATCTT	17820
TCTAGTCTTA TTTCGTCTTA TATTGTAGAC AGCAAGTTGG CTAGTAAGTT CCAAGAAGAG	17880
GTTCAAACAC TTATAAATCA AGAAATAGTA GTTCAAGCTT TTTTGAAGT TATTTGAAGG	17940
ACAGTCCAAT GATGAACACA AACCTGTGTC TTTCsTGGTC TTTTtTAGTG TTTTGAAGGG	18000
TGGKATACTA ATCTCAAAGA TAACAATTAT ATCCAAAGGA GGCAACATAT GCCAAACGTC	18060
AAAGAAATTA CAAGAGAGTC ATGGATTTTA GCCACTTTCC CAGAGTGGGG AACATGGTTG	18120
AACGAAGAAA TCGAAGAAGA AGTCGTACCT GAAGGCAACT TTGCCATGTG GTGGCTAGGC	18180
AACTGTGGTA CTTGGATTAA GACACCAGCT GGTGCTAAGC TTGTCATGGA CCTTTGGTCA	18240
AACCGTGGAA AATCAACCAA AAAAGTGAAA GATATGGTTC GTGGGCACCA AATGGCAAAT	18300
ATGGCAGGTG TTCGTAAGCT GCAACCAAAC TTGCGTGTTC AGCCAATGGT TATCGATCCA	18360
TTTGCTATCA ACGAACTAGA CTATTACTTA GTTTCACACT TCCACAGTGA TCATATCGAC	18420
CCATACACAG CTGCAGCAAT TCTCAATAAT CCTAAGTTAG AGCATGTTAA GTTGG	18475

(2) INFORMATION FOR SEQ ID NO: 39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7186 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

CCAGGATTTG GTACCGTTGC AAGTGGTGTG CCTTTCCTCC TAAAGGAAAA TGGAGGAAAA	50
ATCAATCAAT CAGCACATTC AGATATCAAA GTTGCTAAGG TATTGGTCAA GGATGAAGAT	120
GAAAAAATC GCTTGCTTGC AGCAGGGAAT GACTTTAACT TTGTAACCAA TGTGGATGAT	180
ATTTTATCAG ACCAGGATAT TACTATCGTA GTGGAATTGA TGGGGCGTAT TGAGCCTGCT	240
AAAACCTTTA TCACTCGTGC CTTGGAAGCT GGAAAACACG TTGTTACTGC TAACAAGGAC	300
CTTTTAGCTG TCCATGGCGC AGAATTGCTA GAAATCGCTC AAGCTAACAA GGTAGCACTT	360
TACTACGAAG CAGCAGTTGC TGGTGGGATT CCAATTCTTC GTACTTTAGC AAATTCCTTG	420
GCTTCTGATA AAATTACGCG CGTGCTTGGA GTAGTCAACG GAACTTCCAA CTTTCATGGTG	480
ACCAAGATGG TGGAAGAAGG CTGGTCTTAC GATGATGCTC TTCCGGAAGC ACAACGTCTA	540



386

GGATTTGCAG AAAGCGATCC GACGAATGAC GTAGATGGGA TTGATGCAGC CTACAAGATG	600
GTTATTTTGA GCCAATTTGC CTTTGGCATG AAGATTGCCT TTGATGATGT AGCCCAACAG	660
GGAATCCGCA ATATCACACC AGAAGACGTA GCTGTAGCTC AAGAGCTTGG TTACGTAGTG	720
AAATTGGTTG GTTCTATTGA GGAACTTCT TCAGGTATTG CTGCAGAAGT GACTCCAACC	780
TTCTACCTA AAGCGACCC ACTTGCTAGT GTGAATGGCG TATGAACGC TGTCTTGTGA	840
GAATCTATCG GTATTGGTGA GTCTATGTAC TACGGACCAG GTGCGGTCA AAAACCAACT	900
GCAACAAGTG TTGTAGCTGA TATTGTCCGT ATCGTTCGTC GTTTGAATGA TGGTACTATT	960
GGCAAAGACT TCAACGAATA TAGCCGTGAC TTGTCTTGG CAAATCCTGA AGATGTCAAA	1020
GCAAATACT ATTTCTCAAT CTTGGCTCTA GACTCAAAAG GTCAGGTCTT GAAGTTGGCT	1080
GAAATCTTCA ATGCTCAAGA TATTTCTTTT AAGCAAATCC TTCAAGATGG CAAAGAGGGT	1140
GACAAGGCGC GTGTCGTTAT CATCACACAC AAGATTAATA AAGCCCAGCT TGAAAATGTC	1200
TCAGCTGAAT TGAAGAAGGT TTCAGAATTC GACCTCTTGA ATACCTTCAA GGTGCTAGGA	1260
GAATAAGATG AAGATTATTG TACCTGCAAC CAGTGCCAAT ATCGGGCCAG GTTTTGAATC	1320
GGTCGGTGTA GCTGTAACCA AGTATCTTCA AATTGAGGTC TGCGAAGAAC GAGATGAGTG	1380
GCTGATTGAA CACCAGATTG GCAAATGGAT TCCACATGAC GAGCGTAATC TCTTGCTCAA	1440
AATCGCTTTG CAAATTGTAC CAGACTTGCA ACCAAGACGC TTGAAAATGA CCAGTGATGT	1500
CCCTTTGGCG CGCGGTTTGG GTTCTTCCAG CTCGGTTATC GTTGCTGGGA TTGAACTAGC	1560
CAACCAACTG GGTCAACTCA ACTTATCAGA CCATGAAAAA TTGCAGTTAG CGACCAAGAT	1620
TGAAGGGCAT CCTGACAATG TGGCTCCAGC CATTTATGGT AATCTCGTTA TTGCAAGTTC	1680
TGTTGAAGGG CAAGTCTCTG CTATCGTAGC AGACTTTCCA GAGTGTGATT TTCTAGCTTA	1740
CATTCCAAAC TATGAATTAC GTACTCGCGA CAGCCGTAGT GTCTTGCCTA AAAAAATTGTC	1800
TTATAAGGAA GCTGTTGCTG CAAGTTCTAT CGCCAATGTA GCGGTTGCTG CCTTGTGGC	1860
AGGAGACATG GTGACCGCTG GGCAAGCAAT CGAGGGAGAC CTCTTCCATG AGCGCTATCG	1920
TCAGGACTTG GTAAGAGAAT TTGCGATGAT TAAGCAAGTG ACCAAAGAAA ATGGGGCCTA	1980
TGCAACCTAC CTTTCTGGTG CTGGGCCGAC AGTTATGGTT CTGCCTTCTC ATGACAAGAT	2040
GCCAACAATT AAGGCAGAAT TGGAAAAGCA ACCTTTCAAA GGAAACTGC ATGACTTGAG	2100
AGTTGATACC CAAGGTGTCC GTGTAGAAGC AAAATAAAGA ATAGAAGATA GGATGGGGAA	2160
ACTCTTGACC AGAGGGGTTT ATATCTTTT TGTGAAAAGA AGTTTATACT CAATGAAAAT	2220
CAAAGAGCAA ACTAGGAAGC TAGCCGCAGG CTGCTCAAAA CAGTGTTTTG AGGTTGCAGA	2280
TAGAACTGAC GAAGTCAGCT CAAGACACTG TTTTGAGGTT GCAGATAGAA CTGACGAAGT	2340

387

CAGTAACCAT ACTACGGTAA GGTGACGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTAT	2400
TAGTTAAAAA CGTGATAAAG GAGAAATAAA GATGGCAGAA ATTTATCTAG CAGGTGGTTG	2460
TTTTTGGGGC CTAGAGGAAT ATTTTTCACG CATTTCTGGA GTGCTAGAAA CCAGTGTGG	2520
CTACGCTAAT GGTCAAGTCG AAACGACCAA TTACCAGTTG CTCAAGGAAA CAGACCATGC	2580
AGAAACGGTC CAACTGATTT ACGATGAGAA GGAAGTGTCA CTCAGAGAGA TTTTACTTTA	2640
TTATTTCCGA GTTATCGATC CTCTATCTAT CAATCAACAA GGAATGACC GTGGTCGCCA	2700
ATATCGAACT GGGATTTATT ATCAGGATGA AGCAGATTTG CCAGCTATCT ACACAGTGGT	2760
GCAGGAGCAG GAACGCATGC TGGGTCGAAA GATTGCAGTA GAAGTGGAGC AATTACGCCA	2820
CTACATTCTG GCTGAAGACT ACCACCAAGA CTATCTCAGG AAGAATCCTT CAGGTTACTG	2880
TCATATCGAT GTGACCGATG CTGATAAGCC ATTGATTGAT GCAGCAAACT ATGAAAAGCC	2940
TAGTCAAGAG GTGTTGAAGG CCAGTCTATC TGAAGAGTCT TATCGTGTCA CACAAGAAGC	3000
TGCTACAGAG GCTCCATTTA CCAATGCCTA TGACCAAACC TTTGAAGAGG GGATTTATGT	3060
AGATATTACG ACAGGTGAGC CACTCTTTTT TGCCAAGGAT AAGTTTGCTT CAGGTTGTGG	3120
TTGGCCAAGT TTTAGCCGTC CGATTTCCAA AGAGTTGATT CATTATTACA AGGATCTGAG	3180
CCATGGAATG GAGCGAATTG AAGTTCGTTT TCGTTCAGGC AGTGCTCACT TGGGTCATGT	3240
TTTCACAGAT GGACCGCGGG AGTTAGGCGG CCTCCGTTAC TGTATCAATT CTGCTTCTTT	3300
ACGCTTTGTG GCCAAGGATG AGATGGAAAA AGCAGGATAT GGCTATCTAT TGCCTTACTT	3360
AAACAAATAA AACAGAGAGT GGGGCTTCCC ACTTCTTCA TTTCTAGAAT ATGAATAGAA	3420
GGGATTTATG AAACACCTAT TATCTTACTT CAAACCCTAC ATCAAGGAAT CAATTTTAGC	3480
CCCCTTGTTT AAGCTGTTAG AAGCTGTTTT TGAGCTCTTG GTTCCCATGG TGATTGCTGG	3540
GATTGTTGAC CAATCTTTAC CTCAGGGAGA TCAAGGTCAT CTCTGGATGC AGATTGGCCT	3600
GCTCCTTATC TTTGCAGTAA TTGGCGTTTT AGTGGCCTTG ATAGCTCAAT TTTACTCAGC	3660
AAAGGCAGCA GTAGGTTCTG CTAAGGAATT GACAAACGAT CTTTATCGTC ATATTCTTTC	3720
CTTGCCCAAG GACAGCAGAG ACCGTCTGAC AACTTCTAGT TTGGTCACTC GCTTGACTTC	3780
GGATACCTAC CAGATTCAGA CTGGTATCAA TCAATTCCTG CGTCTCTTTT TACGAGCGCC	3840
CATTATCGTT TTTGGTGCCA TTTTATGGC TTATCGAATC TCAGCTGAGT TGACTTTCTG	3900
GTTCTTAGTC TTGGTTGCCA TTTTGACCAT TGTCATTGTA GGGTTATCTC GATTGGTCAA	3960
TCCTTTCTAC AGTAGTCTCA GAAAGAAAAC GGACCAACTG GTTCAGGAAA CGCGCCAGCA	4020
ATTGCAAGGG ATGCGGGTTA TTCGTGCTTT TGGTCAAGAA AAACGAGAGT TACAGATTTT	4080

388

TCAAACCCCTT AACCAAGTTT ATGCTAGATT ACAAGAAAAG ACAGGTTTCT GGTCTAGTTT	4140
ATTAACACCT CTGACCTATC TGATTGTCAA TGGAAGCTCTT CTCGTTATTA TCTGGCAAGG	4200
CTATATTTCA ATTCAAGGAG GAGTGTCTAG TCAAGGTGCT CTCATTGCTC TTATCAATTA	4260
CCTCTTACAG ATTTTGGTGG AATTGGTCAA GCTAGCCATG TTGATCAATT CCTCAACCA	4320
GTCCTATATC TCAGTCAAGC GAATCGAGGA AGTCTTTGTT GAGGCTCCAG AGGATATCCA	4380
TTCAGAGTTA GAACAAAAGC AAGCTACCAG AGATAAGGTT TTACAAGTCC AAGAATTGAC	4440
CTTTACCTAT CCTGATGCGG CCCAGCCTTC TCTGAGATAC ATTTCTTTTG ATATGACTCA	4500
AGGACAAATT CTAGGTATCA TCGGGGGAAC TGGTTCTGGT AAATCAAGCT TGGTGCAACT	4560
CTTACTTGGA CTTTATCCAG TAGACAAGGG GAACATTGAC CTTTATCAAA ATGGACGTAG	4620
TCCTCTTAAT TTGGAGCAGT GCGGGTCTTG GATTGCCTAT GTACCTCAAA AGGTGCAACT	4680
CTTTAAAGGA ACCATTCGTT CCAACTTGAC TCTAGGTTTC AATCAAGAAG TATCTGACCA	4740
GGAAGCTCTGG CAGGCCTTGG AGATTGCGCA AGCTAAGGAT TTTGTCAAGT AAAAGGAAGG	4800
ACTCTTGGAT GCTCTAGTTG AGGCAGGGGG GCGAAATTTT TCAGGTGGAC AAAAACAAAG	4860
ATTGTCTATC GCCCAGAGCAG TCTTGCGCCA GGCTCCGTTT CTCATCCTAG ATGATGCAAC	4920
CTCGGCACTG GATACCATTA CAGAGTCCAA GCTCTTGAAA GCTATTAGAG AAAATTTTCC	4980
AAACACGAGC TTAATTTTGA TCTCTCAACG AACCTCAACT TTACAGATGG CGGACCAGAT	5040
TCTCCTCTTG GAAAAAGGTG AGTTGCTAGC TGTGGCAAG CACGATGACT TGATGAAATC	5100
CAGCCAAAGC TATTGTGAAA TCAATGCATC CCAACATGGA AAGGAGGACT AGAATGAAAC	5160
GACAAACTGT AAACCAGACG CTCAAACGTT TAGCCGTAGA TTTAGCAAGC CATCCTTTCC	5220
TCCTTTTCCT AGCCTTTCTA GGAAGTATTG CCAAGTTGG CTTATCAATT TACCTACCTA	5280
TTCTGATTGG GCAGGTCATT GACCAAGTCC TAGTGGCTGG TTCATCACCA GTTTTTTGGC	5340
AGATTTTCTT CCAGATGCTC TTGGTGGTAA TAGGAAATAC TCTGGTACAA TGGGCCAATC	5400
CTCTCCTCTA TAATCGTCTA ATCTTCTCTT ATACCAGAGA TTTACGGGAG CGAATCATCC	5460
ATAAGCTCCA TCGTTTACCG ATTGCCTTTG TAGATAGGCA AGGTAGTGGG GAGATGGTTA	5520
GTCGTGTAAC CACGGACATC GAACAGTTGG CAGCTGGCTT GACCATGATT TTAAACCAAT	5580
TTTTTCATTG TGTTTTGATG ATTTTGGTCA GTATTCTAGC CATGCTCCAA ATTCATCTCC	5640
TCATGACTCT CTTAGTCTTG CTGTTGACGC CACTGTCCAT GGTGATTTCG CGCTTTATTG	5700
CCAAGAAATC CTATCATCTC TTCCAGAAGC AAACAGAGAC GAGGGGAATT CAGACTCAGT	5760
TGATTGAAGA ATCGCTTAGT CAGCAGACTA TAATCCAGTC CTTCAATGCT CAAACAGAAT	5820
TTATCCAAAG ATTGCGTGAG GCTCATGACA ACTACTCAGG CTATTCTCAG TCAGCCATCT	5880

389

TTTATTCTTC AACGGTCAAT CCTTCGACTC GCTTTGTAAA TGCACCTATT TATGCCCTTT	5940
TAGCTGGAGT AGGAGCTTAT CGTATCATGA TGGGTTGAGC CTTGACCGTC GGTCTTTAG	6000
TGACTTTTTT GAACTATGTT CAGCAATACA CCAAGCCCTT TAACGATATT TCTTCAGTGC	6060
TAGCTGAGTT GCAAAGTGCT CTGGCTTGGC TAGAGCGTAT CTATGGAGTC TTAGATAGCC	6120
CTGAAGTGGC TGAACAGGT AAGGAAGTCT TGACGACCAG TGACCAAGTT AAGGGAGCTA	6180
TTTCCTTTAA ACATGTCTCT TTTGGCTACC ATCCTGAAAA AATTTTGATT AAGGACTTGT	6240
CTATCGATAT TCCAGCTGGT AGTAAGGTAG CCATCGTTGG TCCGACAGGT GCTGGAAAAT	6300
CAACTCTTAT CAATCTCCTT ATGCGTTTTT ATCCCATTAG CTCGGGAGAT ATCTTGCTGG	6360
ATGGGCAATC CATTTATGAT TATACACGAG TATCATTGAG ACAGCAGTTT GGTATGGTGC	6420
TTCAAGAAAC CTGGCTCACA CAAGGGACCA TTCATGATAA TATTGCCCTT GGCAATCCTG	6480
AAGCCAGTCG AGAGCAAGTA ATTGCTGCTG CCAAAGCAGC TAATGCAGAC TTTTTCATCC	6540
AACAGTTGCC ACAGGGATAC GATACCAAGT TGGAAAATGC TGGAGAATCT CTCTCTGTCC	6600
GCCAAGCTCA GCTCTTGACC ATAGCCCGAG TCTTTCTGGC TATTCCAAAG ATTCTTATCT	6660
TAGACGAGGC AACTTCTTCC ATTGATACAC GGACAGAAGT GCTGGTACAG GATGCCTTTG	6720
CAAAACTCAT GAAGGGCCGC ACAAGTTTCA TCATTGCTCA CCGTTTGTC ACCATTTCAGG	6780
ATGCGGATTT AATTCTTGTC TTAGTAGATG GTGATATTGT TGAATATGGT AACCATCAAG	6840
AACTCATGGA TAGAAAGGT AAGTATTACC AAATGCAAAA AGCTGCGGCT TTAGTTCTG	6900
AATAAGCCAT TCTCTTTTGA AAGTTTATGG ACGAAAAAAG TTGCCTTCGA GTGACTTTTT	6960
TGTTACAATA GCTAGAAAAA TTGTTCACTG TAATACTCAA TGAATATCAA AGAGCAAAC	7020
AGGAAGCTAG CCGTAGGTTG CTCAAAGCAC AGCTTTGAGG TTGTAGATAA GACTGACGAA	7080
GTCAGTTCAA AACACTGTTT TGAGGTTGCA GATAGAAGT ACGAAGTCAG CTCAAAACAC	7140
TGTTTGTAGG TTGCAGATAG AACTGACGAA GTCAGCTCAA AACAGG	7186

(2) INFORMATION FOR SEQ ID NO: 40:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14273 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CTGAAAATTC TAAAAAATTT ATAAGTAAGG AATTAATTAG TTATTTTGT GATAAAGTTT	60
--	----

390

ATGATGAAAT ATTTGTTGAA GAGGTAGTTC CGCACGTTTT TCTGCCATAT GAATCTGACT	120
TACTTCTTAT TTTACCAGCT ACCGCAAATG TGATTGGCAA AATTGCTAAT GGTATTGCTG	180
ATGATTTAGT TACAGCAACT GTTTTAACT TTAATAAAAA AATAATTTTT TGTCCCAATA	240
TGAACCTCTAC TATGTGGGAC AATCACATAG TTCAAAGAAA TGTATCAATT CTAAAGGAGT	300
TGGGACATAT ATTTTATTT GAGTCTAAAA AACATATGA GGTAGGATTG CGTAAAGCAA	360
TAGATTCAAC ATGTTCAATG TTACAACCAC AGTCGTTAGT AAAAGAACTT ATCAAATTAG	420
AAAATATTGT CCTTGAAGAG GGACATTAAA AACTACTGAG AATATTAATG AGGGGAAAAA	480
ATGGAAAATT CATCAATCGA TGTAGATATG CTGTTGGAAG AATTGACACA AGAAGCAATG	540
GTCGTTGTTG CTGTTGATAA GGACTGTAA TTTAACTTA TGGCAATATA TGAAAGGTTA	600
CTGGATGTTT TAAATTATGC AGGCAGTAGC CTTTTATTAT ATACAAATGG ATAAAGTAAG	660
GATAATACAA TGATTAATAA AAAAATACAA CAAGTTGTTT TGGAAATCATT ACAGAATTTT	720
TTGAATGGGA ACTTCATTC GCCTTGTTGA GTCTATGATT TTGGCTTGCT GGAAACTGTA	780
CTTGATGAAT TTAATAATCA AATTCCTGTA ACATTCAATT ACCAACTTTT TTATGCCGTT	840
AAAGCAAAT CAAATGAGAA GATACTTGAA TTCTTAGTAG ATAAATTTGA TGGAGTTGAT	900
GTGGCGTCAT TATCTGAATT AGATGTGGCT AAAAAATTTT TCCCACCAAC TCAAATTTCT	960
GTTAATGGTC CCGCATTTTC TTATGAACT TTATATAATC TGATTAAAAA ACAATATAAA	1020
GTTGATATTA ACTTTTGGGA ACATCTTCAA CAATTTTCCC CAAAAGAATC TGTTGGAATA	1080
AGAGTAACGG AGCCAGATGA ACTTAATAAT CGTATGAGTC GATTTGAAT AAATATTTGC	1140
AGTGATAATT GGACTAGTAA TTTACAAAAT CCTTTAATTA CACGACTGCA TTTTCATTTT	1200
GGAGAAAAAG ATGATAAATT TATGTTAAG TTAGATAAAA TATTATTTAA GTTACAAGAA	1260
ATTAATAAAC TTAGAGAGGT TAGAGAAATA AATCTTGGAG GCGGTTTTAT GAAATTATTT	1320
ATGGAAAATC GTTTGAAAGA ATTTTTTCTA TCACTTATGG AAATCTATAA AAAGTACGAT	1380
ATTGATAGTA CTGTGACTAC AATAATAGAA CCAGGTAGTG CAATTACTTC ATTTTCTGCC	1440
TATATGATTA CTAGCCCAGT TAATGTTAGT GAGGTGAATG AGCAGCAGGT TATCACGTTA	1500
GACACATCAA TATACACCAA TACATTATGG TTTGTTCCGC ATATTATTAC AACGTTAAAT	1560
TCAAGTAGTA AAGAGCGTTA TAGTACTATT CTCTATGGTA ATACCTGTTA TGAACATGAC	1620
AAGTATAAAA TGAAAGTTTC GCTTCCAAGG TTAACCTCAA ATAGCAGTAT AGTGTTTTTT	1680
CCTGTAGGAG CTTATATAAA AAGCAATCAT TCAAATTTAC ATCGTAATGA TTTTATGCGG	1740
GAGGTATATT TGTGGACAAA AAACCTTGACA TATTAGATAA AGTTAAGGAA TATTTAGGAA	1800
ATAAACTAC TCAAATTCG GATAATCAAT ATAAAGAATT TTTGAACTT AATGATATAA	1860

391

GGCGAGCGTT TGGTATTTCA GAAAAAGTAT TAAACAATTC TTTTAATTTT ACGAGTAAAG	1920
AATTTAATGA TTTAATTAAT AACGAAAATT ATTTATTCGA ATATGCATGT AGAATTAGAG	1980
AGGAATGGAG AAAAAAATGC TTTAATCATT CTTATCGTTT TCTATGCTCA CCTATAATTA	2040
CAGATGATTT TCTTAACACG AAGACATTGA GAAGTAGCCA AATTGAATAT AAATATGAGC	2100
GATATTTATC GAAAAGTTTC ATAGGCGATA GAGCGGTTGA TGGCTTTGTT TCCTTCAATA	2160
CTTTAACAGC TAATGGTATG TCTGCTATTA AACTATGTCT TGAGATATTA AACTCTATTT	2220
TCTTCAAGAA GAAGATTGAT TTATTATATT CAACCGGATA TTATGAAACA AGATTTTTAT	2280
TAAATAATCT TGCTAAATCA GGTATTAGTT GCTATGAGGT AAGTAATTGT GAATTGGATA	2340
AAGATAAATT TTATAATGTA TTCATGATGG AACCCAAATCG AGCCGATTTA ACATTACAAA	2400
AAACTGATTT CAAGATAGTA GAATATTTTG TTAAGTATAA AAATAATTCA ATAAAAGTCG	2460
TTATTTTAGA TATTTTCATAT CAAGGTTCTA ATTTTAAATT AGTAGAATTT TTAGAGAAAT	2520
TTAAATTTGC GAATGTAATT ATTTTGTGG TACGATCTTT GATAAAATTA GATCAAATCG	2580
GATTAGAATT GACAAATGGG GGAATAATAG AAGTGTTTAT TCCTAATCAT TTGAGAAAGT	2640
TGAAAAATTT TATGCAAGAG GAATTCAATA AATTTAGAAA TTCTCACGGA GCTAATCTAA	2700
GCCTCTATGA ATACTGTTTG CTGATAAATT CTTAACTTT AAAAAATGAT TGGAACTATT	2760
CTGATTTAGT TATGAAATTT ACGAGTAATT TTTATGCTGA TATAAAAGAC TTGTTTCATGG	2820
AAAATTTCTGA TATGAAATC ATCCATGAAG AGGGAGTACC TTTTGTATTT TTAGATTTAA	2880
TAGGTGAAGG TAAAAAAGAA TATGAAATGT TTTTCAATG GTTAAACTTC TTTTACAAAC	2940
AGCTTGGAAAT CACATTGTAT GCTAGAAAATA GTTTTGGGTT TCGGAATCTA ACAGTAGAGT	3000
ATTTTGGAAAT TATTGGGACA GAAAGATATA TATTTAAGAT TTGTCCAGGT GTTTATAAAG	3060
GGTTAAGTTA TTATTTGATG AAATTTTAT TAAATCTTT TTCAAATGAA TATTTAAAAA	3120
CTACTGATGA GGTTAATAGA TGAAAAATTT GATAAAGTTG CTAATAATTA GATTGATTGT	3180
TAACTTAGCA GACAGTGTAT TTTATATAGT AGCATTGTGG CACGTTAGCA ATAATTATTC	3240
TTGAGCATG TTCTTAGGAA TATTTATTGC AGTAAATTAT CTACCGGATT TGTTACTAAT	3300
CTTTTTTGA CCAGTTATG ACAGAGTAAA TCCGCAAAAA ATTCTTATAA TATCAATTTT	3360
GGTTCAATTA GCAGTGGCTG TAATATTTT ATTATTATTA AACCAAATAT CATTTTGGGT	3420
GATAATGAGT CTAGTGTTA TTTCAGTAAT GGCTAGCTCC ATAAGTTACG TGATAGAAGA	3480
TGTGTTGATT CCTCAAGTGG TAGAATATGA TAAGATTGTA TTGCAAAAT CTCTTTTATG	3540
TATTTCTGAT AAAGTATTAG ATTCTATTTT TAATTCATTC GCATCATTTT TACAGGTGGC	3600

392

AGTAGGATTT ATTTTATTGG TTAAGATAGA TATAGGCATA TTTTACTTG CTCTATTTAT	3660
ATTGTTGTTG TTAATAATTA GAACTAGCAA TCGGAATATA GAAAACCTCT CTTTCAAATA	3720
TTACAAGAGA GAAGTGTTC AAGGTACAAA GTTTATTTTA AATAATAAAT TATTATTTAA	3780
AACCAGTATT TCTTTAACGC TTATAAACTT TTTTATTCA TTTCAGACAG TAGTTGTACC	3840
GATTTTTTCT ATTCGATATT TTGATGGTCC GATTTTTTAT GGTATTTTTT TAACTATTGC	3900
TGGTTTGGGT GGTATATTGG GAAATATGCT AGCGCCAATC GTAATAAAAT ATTTAAAATC	3960
GAATCAAATT GTTGGTGTAT TTCTTTTTTT GAACGGCTCA AGTTGGTTAG TAGCAATTGT	4020
TATAAAGAC TATACTTTAT CACTTATTTT ATTTTTCGTT TGTTTTATGT CTAAGGAGT	4080
CTTCAATATT ATTTTAAAT CGTTGTACCA ACAAATACCT CCACATCAAC TTCTGGTAG	4140
GGTAAATACT ACCATTGATT CTATTATTTT TTTTGGAAAT CCAATTGGTA GTTTAGTTGC	4200
AGGAACGCTT ATTGATTGCA ATATTGAATT AGTGTTAATT GCTATTAGCA TACCTTATTT	4260
TTTGTTTCT TATATTTTTT ATACGGATAA TGGATTGAAA GAATTTAGTA TATATTAGAA	4320
ATGTTTATGT TCATTCAAAA GCATAATGAC TATAACTGAA AAAGAAAAGT GATATCTTTA	4380
AGGTTGTCT TCTTGGTGGT GAGATTCGTG AGACAACCCA AGCTTTTCTC GGAAAGATTA	4440
CCAATGCTTT GATGGATAGG ATGTACTTTA GCAAGATGTT TTTAGTCGTA ACGGTATCGT	4500
GGATGGACGT GTAATAACCT CTTCTTTTGA GGAGTATTTT ACTAAAAAAC TAGCCTTGA	4560
GCGTTCCCA GAAACGGACT TACTCATTGA CTCTTCAAAG ATTTGGGGAG AAGATTTTGC	4620
TTCATCTGTT CCTTGAAAA AGTCACAGCA GTCATCACAG ACGATAGTAC TGAACAAAAC	4680
TATGAAGAGT TAGAAATTA TACCGAGGTG ATTGTATAAA GGATCTGGAA ATAGATAAGA	4740
AGTTGATTAG TATTGACCTA GGTGGTACAA ATATTAAGAT TACTGTCTCT TCAAATGACG	4800
GTGAGATTGA AACTTTGTGG AGTATTACAA CAGATACAAG TGAGAAAGGT TCTCAAATTA	4860
TATCGGACAT CATCAGTTCT ATTAATAATA AATTGACCGA ACGGAATATT CCTGATAGCG	4920
ACCTTCTTGG AATCGGTATG GGAAGTTGCT CATCATACTT TCCTTGTAAA TCATAGGGGC	4980
TATAAACTCT CCGTCTACTT GTCCTGCAAC AATTGAAGTC TGCTCAAAAC GCCGTCCGCT	5040
AATCTTTTCA TAGACTTTCT CCCTTTTAGG AGCCTAGCTT TCTAGTTTGT TCTTTGATTT	5100
TTATTGAGTA TACCACTATT TTAATCCCTC TGGCAAGGGA CTTTGTCTAT GTGGAGGGAT	5160
TGGGCTCCTA TGTGGTGGAG CTTTCTGTCT CTTTCTGAAA TATGGTATAA TAGCACTAAT	5220
CAATTTCTAG GAAAATAGAT ACAGAAAGGG GCTGAAAGAT GTCTCATATT ATTGAATTGC	5280
CAGAGATGCT GGCAACCAA ATCGCGGCTG GAGAGGTCAT TGAACGTCCT GCCAGTGTGG	5340
TCAAAGAGTT GGTAGAAAAT GCCATTGACG CGGGCTCTAG TCAGATTATC ATTGAGATTG	5400

393

AGGAAGCTGG TCTCAAGAAG GTTCAAATCA CGGATAACGG TCATGGAATT GCCCAGCATG	5460
AGGTGGAGTT GGCCCTGCGT CGCCATGCCA CCAGTAAGAT AAAAAATCAA GCAGATCTCT	5520
TTCGGATTCC GACGCTTGGT TTTCGTGGTG AAGCCTTCCC TTCTATTGCG TCTGTTAGTG	5580
TCTTGACTCT GTTAACGGCG GTGGATGGTG CTAGTCATGG AACCAAGTTA GTCGCGCGTG	5640
GGGGTGAAGT TGAGGAAGTC ATCCCAGCGA CTAGTCCTGT GGGAAACCAAG GTTTGTGTGG	5700
AGGATCTCTT TTTCAACACG CCTGCCCCGC TCAAGTATAT GAAGAGCCAG CAAGCGGAGT	5760
TGTCTCATAT CATTGATATT GTCAACCGTC TGGGCTTGGC CCATCCTGAG ATTTCTTTTA	5820
GCTTGATTAG TGATGGCAAG GAAATGACGC GGACAGCAGG SACTSGTCAA TTGCGCCAAG	5880
CAATCGCAGG GATTACGGT TTGGTCAGTG CCAAGAAGAT GATTGAAATT GAGAATCTG	5940
ACCTAGATTT CGAAATTTCA GGTTTTGTGT CCTTGCCTGA GTTGACTCGG GCTAACCGCA	6000
ATTATATCAG CCTCTTCATC AATGGCCGTT ATATTAAGAA CTTCCTGCTC AATCGTGCTA	6060
TTTTGGATGG TTTTGAAGC AAGCTTATGG TTGGACGTTT TCCACTGGCT GTCATTACACA	6120
TCCATATCGA CCCTTATCTA GCGGATGTCA ATGTGCATCC AACTAAGCAA GAGGTGCGGA	6180
TTTCCAAGGA AAAAGAACTG ATGACTCTGG TTTCAGAAGC TATTGCAAAT AGTCTCAAGG	6240
AACAAACCTT GATTCCAGAT GCCTTGGAAG ATCTTGCCAA ATCGACCGTG CGCAATCGTG	6300
AGAAGGTGGA GCAAACTATT CTCCCACTCA AAGAAAATAC GCTCTACTAT GAGAAAAGTG	6360
AGCCGTCAAG ACCTAGTCAA ACTGAAGTAG CTCATTATCA GGTAGAATTG ACTGATGAAG	6420
GGCAGGATTT GACCCTGTTT GCCAAGGAAA CCTTGGACCG ATTGACCAAG CCAGCAAAAC	6480
TGCATTTTGC AGAGAGAAAG CCTGCTAACT ACGACCAGCT AGACCATCCA GAGTTAGATC	6540
TTCTAGCAT CGATAAGGCT TATGACAAAC TGGAGCGAGA AGAAGCATCC AGCTTCCCAG	6600
AGTTGGAGTT TTTCGGACAA ATGCACGGGA CTTATCTCTT TGCCCAAGGG CGAGATGGAC	6660
TTTACATCAT AGATCAGCAC GCTGCTCAGG AACGGGTCAA GTACGAGGAG TACCGTGAAA	6720
GCATTGGCAA TGTTGACCAA AGCCAGCAGC AACTCCTAGT GCCCTATATC TTTGAATTTT	6780
CTGCGGATGA TGCCCTGCGT CTCAAGGAAA GAATGCCTCT CTTAGAGGAA GTGGGCGTCT	6840
TTCTAGCAGA GTACGGAGAA AATCAATTTA TTCTACGTGA ACATCCTATT TGGATGGCAG	6900
AAGAAGAGAT TGAATCAGGC ATCTATGAGA TGTGCGACAT GCTCCTTTTG ACCAAGGAAG	6960
TTTCTATCAA GAAATACCGA GCAGAGCTGG CTATCATGAT GTCTTGCAAG CGATCTATCA	7020
AGGCCAATCA TCGTATTGAT GATCATTCAG CTAGACAACCT CCTCTATCAG CTTTCTCAAT	7080
GTGACAAATCC CTATAACTGT CCTCACGGAC GTCCGTGTTT GGTGCATTTT ACCAAGTCGG	7140



394

ATATGGAAAA GATGTTCCGA CGTATTCAGG AAAATCACAC CAGTCTCCGT GAGTTGGGGA	7200
AATATTAAAA GTATAAAAAA GTCTGGGAAA AATTTTCAAA ATCAAAAAAA CGCATAAAAT	7260
CAGGTGTTC AAAAACTTGA TTTTATGCGT TTTATCATGG AAATAGTTAC TTCATTTTTT	7320
CCTAATTCTT TCGAAACTC TTTTAAACG ACGTCAGTTT TATCAGTAAT CTCAAAAACAG	7380
TGTTTTGAGC TAATTTTGCC AGTTTGTCT GTAACATCGA AGTTGTGTTT TACCACTCTG	7440
CGACTGGTTT CCTAGTTTGC TCTATGATT TCACAGAGCA TTAAATTGCG ATTTTGCCAA	7500
GTTTCTTTAT TCGTCTAAAA GTAGAGTCTG TTCTATGCGT CTAATGTACG AATCAGGTTG	7560
ACCATTTCAA TAGCTCCTTG TGCACACTCA GAACCTTAT TTCTGCTTT AGTACCAGCT	7620
CGTTCTATGG CTGTTC AAT TGTATCTGTC GTTAGCACAC CAAACATAAC AGGAATTTCCG	7680
CTATTTAAAC TGATTTGGGC GATTCCTTA GATACCTCGC TACATACATA ATCATAATGA	7740
CTTGATTTCC CTCTAATGAC AGCTCCCAAG CAGATAATTG CATCATATTT TTTACTTTTT	7800
GCCATTTTTG ATGCAATCAG TGGTATTTCA AAAGCTCCTG GAACCCAGGC TACCTCTATA	7860
TCTTTCTCGT TTACATTCTC TCTTTTGAGA TTATCTAGTG CTCCAGATAA TAATTTTGAA	7920
GTTATAAATT CATTAAATCT CGGTACAACA ATACCTATTT TAATATTGTT TGCTACTAAA	7980
TTACCTTCAT AAGTGTTTAT TTATTTTCC TCCATATTTA AAATGTGACC CATTCGATTT	8040
TTCTTTGTTT CTAAATAAAA ACTATCGTAA GGATTGGCTT CTATTTGAT TGATATTCTA	8100
CTGGAAATGG TAATTCATA TTTTCTAAC GTTCAACCT TGTCAGGATT ATTTGTCAGT	8160
AAATGAAGTG ACTGAAGTCC CAGATCTTTA AGCATTTTTG CTCCAATATG ATATTCTCTT	8220
AAATCACCTT CAAAGCCTAA TGCAAGATTG GCATCAAGCG TATCCATGCC TTGATCTTGT	8280
AAATGATAGG CTTTAAATTT ATTGATAAGT CCAATTCCTC GTCCCTCCTG TCGCAAGTAA	8340
AGTAAGACAC CCGAACCATT CTCAACAATC ATTTTCATAG CCTTATCGAA TTGCTGTCCA	8400
CAATCGCAAC GTAAAGAGCC TAAACATCT CCTGTAAAC ATTCGGAGTG GACCCGACAT	8460
AATACATTGG CTTCATCCTC TATAATTTCC ATAATAAGAG CAAGATGATG TTCCCCATTT	8520
AGTTTATCTA TATAGCTAAT TGCTTTGAAA TTACCGTATC TAGTAGGCAT ATTGACAGTT	8580
GAAACTCGTT CTACCAGCTG ATCATATACT TTTCTATATT CTTGTAATTC TTTGATGGTA	8640
ATTAGTGGAA TGTGTGTTTT TTTCGAGAAC TGAATTAAT CATCTGTTCT CATCATTTTG	8700
CCATCATGAT TCATTATTC ACAACATAGG CCACACTCTT TTAGTCCAGC TAATTTTAAT	8760
AAATCAACAG TTGCTTCTGT GTGTCCATTT CTTTCTAGGA CACCACCTTT TTTTGCAATT	8820
AAAGGAAACA TGTGTCCTGG CCTGCGAAAA TCAGAGGGTG TTATATCTTC AGCTACACAC	8880
ATACGTGCGG TCAGTCTCTT TTCTCGGCA GAAATACCTG TGGTCGTTTC TTTATAATCA	8940

395

ATTGAACTG TAAAAGCAGT CTTATGATTA TCTGTATTGT TTTCAACCAT AGGTGAAAGC	9000
ATTAATTGAT TAGCTAAACT TTCGCTCATA GGCATACAAA TTAATCCTTT GGCATAAGTA	9060
GCCATAAAAT TAACATTTTC TGTGTAGCT GCTTGTGCAG AACAAATTAA GTCTCCTTCA	9120
TTTTCTCTAT CCTGTGTCGTC TATAACAAGA ACAAGTCGTC CCTTCTGCAA TGCTTCTAAT	9180
GCTTCTTGTA TTTTTCGATA TTCCATTGAC TGATTATCCT TTCTGCTAAA ATCCATTTTG	9240
ATATAATAGT TCCTTAGATA TTTCTGATTT TGGAGAGTTA TCCATCACTT TTTGCACATA	9300
TTTACCTAAG ATATCATTTT CAAGATTAC TGTACTCCCG ACTTGTTTAC TCTTAAGAAT	9360
GGTTTGTTC AAGGTATGAG GGATAACAGA TACTGAAAAG TTTACTTTGG AGACTTTAGC	9420
GACAGTCAGA CTAATGCCGT CAATTGTAAT AGATCCTTTT TCAACTATTA AATCTAAAAT	9480
TTCTTTTTGT GTGTTGATTT GATACCATAC AGCATTATCA TCTTTTTTTA TTGACGAGAT	9540
TTTTCTGTGA CCATCAATGT GTCCTGTAAC GACGTGACCC CCAAGTCGAC CGTTGACAGA	9600
TAAGGCTCTT TCTAGATTCA CCTCACTTCC ATGTTTTAAT AGAGTAAGAG CTGTTGACT	9660
CCATGTTTCA TTCATTACAT CAACTGTAAA GGATTGATGA TTGAAATGAG TAACTGTAAG	9720
ACAGATACCA TTTACTGCTA TACTATCGCC TAAATGGATA TCCGTTAATA TTTTGTAGGC	9780
TTTAATTGAT AGTTTACAAT TACGAGAGTC TTTCTGTATT CTTTCAACTT TTCCGATTTT	9840
TTCAATTATT CCTGTGAACA TGGATAAATC ACTTCACTTT CTATGAGATA GTCATTTCTT	9900
ATTTGAGAAA ATGCATAAGG TTTCAATCTA ATAGCGTCAT TTGGCAAAGA AATACCTTCA	9960
CCTCCGACAG GAAACTTGGC ACTACCTCCA AAAACTTTTG GTGCAATATA TATTTTCAGC	10020
TCATCAACAA TTTGTGTTT CAAAGCACTC CAATTCATTA GACTGCCCCC TTCTAGAACT	10080
AGGCTATCAA TCTGCATGTT TCCTAGATGT TGCATTAAAC TCGATAAGTC TATATGATTG	10140
CCTTTTTTCT TTATGGAAAG TATTTACAG CCATGATTTT GATATAGCTT CATTTTATTT	10200
TTGTCTTCAG AGGAAGTGGC AATGTAAGTT TTAATATCAT TTGCTGTTTT TACGATTTTA	10260
GAGGTAAGAG GAGTTCGTAA ATGTGTATCG CATATGATAC GGATAGGATT TTTCCCTTCC	10320
TCCAATCTAC ATGTCAGCAA AGGATCGTCT TGAATAACAG TATTGACTCC CACCATAATT	10380
GCACTAACAT GGTGTCGTAA CTGATGCACA TGCTTTCTTG CTTCTTCTTC AGTAATCCAT	10440
TTGGATTGAT TTGTTTTAGT GGCTATTTTT CCATCCATTG ACATTGCATA TTTCATAAAA	10500
ACATAGGGTA CATGCTGGGT AATATACTTT CTAAAACTTT TTATTAAGTT AAGACACTCA	10560
TTTTCTAAAA TTCCAACAGT AACTTGAAGA TTATTTTCCT CAAGTATCTT TACTCCTTTT	10620
CCAGATACAA TAGGATTACA GTCTAGGCTT CCAATGACTA CTCTTGTAAT ACCACTATCG	10680

196

ATTATAGCAT CTATACAGGG AGGTGTTTTT CCGAAGTGAC AACAGGGTTC AAGTGTTACA	10740
TAAAGCGTCG CTCCGACAGG GGATTCTCTA CAGTTTTTAA GAGCATTTCCT CTCAGCATGT	10800
GGGCCACCAA AAAACTCATG ATAACCTTGT CCGATAATGT GATTATCTTT TACAATAACT	10860
GCGCCGACCA TAGGATTGGG ATTGACGTAA CCAGCCCCTT TTTGTGCCAG TTTTATTGCT	10920
AATTTCATAT ATTTTGAATC GCTCATCTCG CTACCTCCAA AAAAATATAC CTTGAATAGG	10980
GGACTACTCA AGGCATACAA AAGAAAACCT ATGCGATTAA CAAAAATGCT CTGAAATGAC	11040
AAGTAATCAT TTCAGAGCAC GCAAAAAGCA CAAATATACT TTTATCTTCT TTCATCCAGA	11100
CTATACTGTC GGCTTTGGAA TTTACCCAAA TCATGCCCTT CGGCTCGTGG GCTATACCAC	11160
CGGTAGGGAA TTTACCCCTG CCTGAAGAT AGTTATTCAA TTACAGATGA TTATAGTACT	11220
TAATTTTGAA TATGTCAACA GATAAATACC GATTGTTTTT GATATACTGT ATTTGTGATA	11280
ATCGATTCTC GCTCCTCGGA TAAAGAAAAT ATGATATACT AGATAAACGA AATAAGAGAG	11340
AAGGAATACT ATGTACGCAT ATTTAAAAGG AATCATTACC AAAATTACTG CCAAATACAT	11400
TGTTCTTGAA ACCAATGGTA TTGGTTATAT CCTGCATGTG GCCAATCCTT ATGCCTATTC	11460
AGGTCAGGTT AATCAGGAGG CTCAGATTTA TGTGCATCAG GTTGTGCGTG AGGACGCCCA	11520
TTTGCTTTAT GGATTTGCTT CAGAGGATGA GAAAAAGCTC TTTCTTAGTC TGATTTGGGT	11580
CTCTGGGATT GGTCTGTAT CAGCTCTTGC TATTATCGCT GCTGATGACA ATGCTGGCTT	11640
GGTCAAGCC ATTGAAACCA AGAACATCAC CTACTTGACC AAGTTCCCTA AAATTGGCAA	11700
GAAAACAGCC CAGCAGATGG TGCTGGACTT GGAAGGCAAG GTAGTAGTTG CAGGAGATGA	11760
CCTTCCTGCC AAGCTCGCAG TGCAAGCAAG TGCTGAAAAC CAAGAATTGG AAGAAGCTAT	11820
GGAAGCCATG TTGGCTCTGG GCTACAAGGC AACAGAGCTC AAGAAAATCA AGAAATTCTT	11880
TGAAGGAACG ACAGATACAG CTGAGAACTA TATCAAGTCG GCCCTTAAAA TGTGGTCAA	11940
ATAGGAGCAG AGAATGACAA AACGTTGTTC GTGGGTCAAG ATGACCAACC CGCTCTACAT	12000
CGCCTATCAT GATGAGGAGT GGGGCCAGCC CCTCCATGAT GACCAAGTAT TGTTTGAGTT	12060
GTTGTGTATG GAAACCTATC AGGCAGGCCT GTCTTGGGAA ACGGTACTCA ACAAACGCCA	12120
AGCTTTCCGA GAAGTCTTTC ATAGCTATCA AATTCATCTA GTCGCAGAGA TGA CTGACAC	12180
TGAATTGGAA GCCATGCTGG AGAATCCAGC TATCATTGCA AATAGAGCCA AGCTTTTTCG	12240
TACACGCGCT AACGCCCAAG CCTTTCTACA GTTACAGGCA GAGTACGGCT CTTTGTATGC	12300
CTATCTTTGG TCTTTTGTG AGGGGAAAAC TGTCGTTAAC GATGTTCTCT ATTATCGCCA	12360
AGCGCCAGCT AAAACACCCT TATCTGAGAA ATTAGCCAAA GATCTCAAAA AACGAGGCTT	12420
CAAGTTCACA GGCCAGTCG CCGTATTGTC TTTTCTACAG GCTGCAGGGC TAGTTGATGA	12480

397

CCACGAGAAT GATTGTGAGT GGAAAGGTCT TAAATGATGT CTAACAAAAA TAAGGAAATT	12540
CTGATTTTGG CGATTCTCTA TACAGTCCTC TTTATGTTTG ATGGCGTTAA ATTGCTGGCT	12600
TCTTTAATGC CATCTGCCAT TGCAAATTAT CTGTTTATG TAGTTTTAGC TCTATATGGC	12660
TCCTTCTTGT TCAAGGATAG ATTGATCCAA CAATGGAAGG AGATTAGAAA GACTAAAAGA	12720
AAATTCTTCT TTGGAGTCTT AACAGGATGG CTCTTTCTCA TTCTGATGAC TGTGTCTTT	12780
GAATTGTAT CAGAGATGTT GAAGCAGTTT GTGGGACTAG ATGGACAAGG TCTAAATCAG	12840
TCTAATATTC AAAGTACCTT TCAAGAACAA CCACTACTGA TAGCTGTTTT TGCTGTGTCT	12900
ATTGGACCTC TGGTAGAAGA ATTATTTTTT CGTCAGGTCT TATTGCATTA CTTCAGGAA	12960
CGGTGTCTAG GTTTACTAAG CATTATTCTG GTAGGACTTG TTTTGTCTCT GACTCATATG	13020
CACAGTTTGG CTCTATCAGA GTGGATTGGT GCAGTTGGTT ACTTAGGTGG AGGCCTTGCC	13080
TTTTCTATTA TTTATGTGAA AGAAAAAGAG AATATCTACT ATCCCTACT TGTTCACATG	13140
TTAAGCAACA GCCTCTCCTT AATCATTTTA GCTATCAGTA TAGTAAATG AAATGAGAAC	13200
AGGACAAATC GATTTCTAAC AATGTTTTAG AAGTAGAGGT GTACTATTCT AGTTTCAATA	13260
TACTGTAATA TGTGATGAAA ATGCCAGTAA TGATACCGAG AAAAAAGCTG AGAAACTTTT	13320
CCCAGCTTTA TTTGTTATAG TCAAAGAGAA TGACTTGTTC CTGTGCATCT ACATGAGCAT	13380
GGACCCCAAA GGGTACAATT GCTCTTGGAG TTGCGTGGCC GACATTCAGA TTATAGACAA	13440
TCGGGATATT GCTGTCAATG ATATCCAATA GTGCCTCTTT ATAGTCGTCA TGGAAAGTTT	13500
CATCCATAGG TTTCCGACC AAGAGTCCAT TGATGACCGC GAATATGCCA GTGTCCTTTA	13560
AAGTTAGCAA CATCTTTTGG AAGTCTTCTG GCTTAGGCTT TTCTTCGCTT GTTTCGAGCA	13620
AGAGGATTTT CCCTTCCGAG TCTGACAGT CAGGGAAAAG TCTGTATCTT TGGCAGGTT	13680
CCGTGCTATC TCGGTATCGA GAGTTGTCAA AGATATCGTA GAGGGATTCT AGGCAACCAC	13740
CGAGGATTTT CCCCTCGAAC TGGGCACTTC CTGCAACAA GTCAAAACCT GTATTTGTAT	13800
GACTGACACG AGGTGTTCCC AGGGCCGTGG GACTAAAATC AGTTCGTTC TCATACCAAA	13860
CGTCACTAGG GCGGATTCTT GAAATTCTTC CCGTCTCAAT CAATTCTTTA AAGTAGTGAA	13920
GGCTATAGGC TAGCATTTCT TTGTCTAATT CACAAATGTC TGCTAAAAAG GATTGACCAT	13980
AAAAAGTCTT GATTCCTAAT TTATGCAACA TGAGGTGGTT CATGGTTGTA TCCGAGAAGC	14040
CAAGAAAAAT TTTTGTCTTG ATAACCTTTT GGAGTTGGTC ATTTTCAAAA AGATAAGGTA	14100
GCAAGCGATA GGTATCGTCT CCACCGATGG CACATAGGAT CATGTGATG CTATCATCAG	14160
AAAAGGCATG AATCAAATCC TCTGCACGAG CTTCAGGATG GTCCTTGATA AAGTCTAATC	14220

398

CTTTTAACGA ATGGGGCAAA AAGATGGGAT TGGTCCCAGA TCCTTGAGAC GTT

14273

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9828 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GTGAAGTGGC GCAAAAGGTG CAAGTGATGA GCTCAGGTTT TTAGCTCTT GACATTGCCC	60
TTGGCTCAGG TGGTTATCCT AAGGGACGTA TCATCGAAAT CTATGGCCCA GAGTCATCTG	120
GTAAGACAAC GGTGGCCCTT CATGCAGTTG CACAAGCGCA AAAAGAAGGT GGGATTGCTG	180
CCTTTATCGA TCGGAACAT GCCCTTGATC CAGCTTATGC TCGGGCCCTT GGTGTCAATA	240
TTGACGAATT GCTCTTGCTT CAACCAGACT CAGGAGAGCA AGGTCTTGAG ATTGCGGGAA	300
AATTGATTGA CTCAGGTGCA GTTGATCTTG TCGTAGTCGA CTCAGTTGCT GCCCTTGTTT	360
CTCGTGCGGA AATTGATGGA GATATCGGAG ATAGCCATGT TGGTTTGCGAG GCTCGTATGA	420
TGAGCCAGGC CATGCGTAAA CTTGGCGCCT CTATCAATAA AACCAAAACA ATTGCCATTT	480
TTATCAACCA ATTGCGTGAA AAAGTTGGAG TGATGTTTGG AAATCCAGAA ACAACACCGG	540
GCGGACGTGC TTTGAAATTC TATGCTTCAG TCCGCTTGGG TGTTCTGGT AATACACAAA	600
TTAAGGGAAC TGGTGACCAA AAAGAAACCA ATGTCCGTAA AGAACTAAG ATTAAGGTTG	660
TAAAAAATAA GGTAGCTCCA CCGTTTAAGG AAGCCGTAGT TGAAATTATG TACGGAGAAG	720
CAATTTCTAA GACTGGTGAG CTTTGAAGA TTGCAAGCGA TTTGGATATT ATCAAAAAAG	780
CAGGGGCTTG GTATTCTTAC AAAGATGAAA AAATTGGCA AGGTTCTGAG AATGCTAAGA	840
AATACTTGGC AGAGCACCCA GAAATCTTTG ATGAAATTGA TAAGCAAGTC CGTTCTAAAT	900
TTGGCTTGAT TGATGGAGAA GAAGTTTCAG AACAAGATAC TGAAAACAAA AAAGATGAGC	960
CAAAGAAAGA AGAAGCAGTG AATGAAGAAG TTCCGCTTGA CTTAGGCGAT GAACTTGAAA	1020
TCGAAATTGA AGAATAAGCT GTTAAAGCAG TGGAGAAATC CGTACTTTT TCGATTTTGT	1080
ATTCAAGTTT TTAGATTATA TATAGTAGCT TGAAATAAGA TATGAACAAC TCTATTAGGA	1140
AAGTCAAATT AATTCTAGA AATGTTTGTG CAGCTACAGC GTACTATTCC AAACCAACC	1200
AACTATAATA GATCGAAACT AGAATAGTAC ATATCTACTT CTAACACATT GTTAAAAATC	1260
GATTTGACTT TCCTTATTTT ATTCCGCTAT ATATAGTTTG CTGTTTCTTG TCGCTCCTCT	1320
GGAAAGCTGA TATAATAGCT TTATGAATAA AAAACGAACA GTGGACCTGA TACATGGTCC	1380

GATTCTTCCC TCGCTCTTAA GCTTCACCTT TCCAATTTTG CTATCAAATA TTTTCAACA	1440
GCTCTATAAC ACTGCTGATG TCTTGATTGT TGGACGATTT CTTGGTCAAG AATCCTTGGC	1500
TGCAGTAGGA GCGACGACAG CGATTTTGA CTTGATTGTA GGTTTTACAC TTGGTGTGG	1560
CAATGGCATG GGGATTGTCA TTGCTCGTTA TTATGGGGCT CGGAATTTC CTAAATCAA	1620
GGAAGCAGTA GCAGCCACCT GGATTTTAGG TGCTCTTTG AGCATTTCTAG TTATGTTGCT	1680
GGGCTTTCTT GGCTTCTATC CTCTCTTGCA ATACTTAGAT ACTCCTGCAG AAATTCTTCC	1740
TCAATCTTAT CAATATATT CTATGATTGT GACCTGTGTA GGTGTCAGCT TTGCTTATAA	1800
TCTTTTGA GGCTTGTGC GGTCTATTGG TGACAGTCTA GCAGCCCTGG GATTCTGAT	1860
TTTCTCTGCC TTGGTTAATG TGGTCTGGA TCTCTATTT ATTACGCAAT TGCATCTGGG	1920
AGTTCAATCC GCAGGACTTG CTACCATTAT TTCGCAAGGT TTATCAGCGG TTCTCTGCTT	1980
TTATTATATT CGTAAAGTG TGCCAGAAGT CTTGCCACAG TTAAACATT TCAAATGGGA	2040
CAAAAGCTTG TACGCGGATC TCTTGGAGCA AGGTTTGGCT ATGGGCTTGA TGAGTTCAAT	2100
TGTATCTATC GGCAGTGTGA TTTTACAGTT TTCTGTTAAT ACATTTGGTG CAGTGATTAT	2160
TAGTGCCAG ACGGCAGCTC GACGCATTAT GACCTTTGCC CTTCTTCTTA TGACCGCTAT	2220
TTCTGCATCA ATGACGACCT TTGCTTCTCA GAATCTAGGA GCTAAGCGAC CTGACCGTAT	2280
TGTTCAAGGT CTTCGAATCG GCAGTCGTTT AAGTATATCC TGGGCAGTTT TTGTTTGTAT	2340
TTTCTCTTT TTGCCAGTC CAGCTTTGGT TTCCTTCTTG GCTAGTTCTGA CAGATGGTTA	2400
CTTGATAGAA AATGGAAGTC TCTATCTGCA AATCAGTTCA ACCTTTTATC CCATTTTGAG	2460
CCTCTGTTG ATTTATCGCA ATTGCTTGCA GGGCTTGGG CAAAAGATCC TTCCTCTAGT	2520
TTCTAGCTTT ATTGAAGTAA TCGGAAAAAT CTTTTTCTG GTTCTTATTA TTCCTTGGGC	2580
AGGATATAAG GGTGTTATCC TTGTGTAACC TCTTATCTGG GTTGCCATGA CAGTTCAACT	2640
GTACTTCTCA TTATTCCGTC ATCCCTTGAT AAAAGAAGGC AAGGCAATCT TGGCAACCAA	2700
AGTGCAATCC TACTTGGATT TACTGAATAA AATCCATTC CTCTAGTGAA AATCGAAAA	2760
ACTTGTGTT TCTTCTTTAG TTTGGTGTG AAAATAGTTT AACAGACTTT TGACTTCTTT	2820
TATATGATAT AATAAAGTAT AGTATTTATG AAAAGGACAT ATAGAGACTG TAAAAATATA	2880
CTTTTGAAAA TCTTTTATAG CTGGGGTGTT ATTGTAGATA GAATGCAGAC CTTGTCAGTC	2940
CTATTTACAG TGTCAAAAA GTGCGTTTTG AAGTTCTATC TACAAGCCTA ATCGTGACTA	3000
AGATTGTCTT CTTTGTAAGG TAGAAATAAA GGAGTTTCTG GTTCTGGATT GTAAAAAATG	3060
AGTTGTTTTA ATTGATAAGC ACTAGAATAT GGAAATTAAT GTGAGTAAAT TAAGAACAGA	3120

400

TTTGCCCTCAA GTCGGCGTGC AACCATATAG GCAAGTACAC GCACACTCAA CTGGGAATCC	3180
GCATTCAACC GTACAGAATG AAGCGGATTA TCACTGGCGG AAAGACCCAG AATTAGGTTT	3240
TTTCTCGCAC ATTGTTGGGA ACGGTTGCAT CATGCAGGTA GGACCTGTTG ATAATGGTGC	3300
CTGGGACGTT GGGGGCGGTT GGAATGCTGA GACCTATGCA GCGGTTGAAC TGATTGAAAG	3360
CCATTCAACC AAAGAAGAGT TCATGACGGA CTACCGCCTT TATATCGAAC TCTTACGCAA	3420
TCTAGCAGAT GAAGCAGGTT TGCCGAAAAC GCTTGATACA GGGAGTTTAG CTGGAATTAA	3480
AACGCACGAG TATTGCACGA ATAACCAACC AAACAACCAC TCAGACCACG TTGACCCTTA	3540
TCCATATCTT GCTAAATGGG GCATTAGCCG TGAGCAGTTT AAGCATGATA TTGAGAACGG	3600
CTTGACGATT GAAACAGGCT GGCAGAAGAA TGACACTGGC TACTGGTACG TACATTCAGA	3660
CGGCTCTTAT CCAAAAGACA AGTTTGAGAA AATCAATGGC ACTTGGTACT ACTTTGACAG	3720
TTCAGGCTAT ATGCTTGCAG ACCGCTGGAG GAAGCACACA GACGGCAACT GGTACTGGTT	3780
CGACAACTCA GCGGAAATGG CTACAGGCTG GAAGAAAATC GCTGATAAGT GGTACTATTT	3840
CAACGAAGAA GGTGCCATGA AGACAGGCTG GGTCAAGTAC AAGGACACTT GGTACTACTT	3900
AGACGCTAAA GAAGGCGCCA TGGTATCAAA TGCCTTTATC CAGTCAGCGG ACGGAACAGG	3960
CTGGTACTAC CTCAAACCAG ACGGAACACT GGCAGACAAG CCAGAATTCA CAGTAGAGCC	4020
AGATGGCTTG ATTACAGTAA AATAATAATG GAATGTCTTT CAAATCAGAA CAGCGCATAT	4080
TATTAGGTCT TGAAAAAGCT TAATAGTATG CGTTTTCTTG TGGAGATATT TCCTTCAATT	4140
TTGCTACTAT ATTAACAAA AATCAAAAAG CAACTAGAA ACTTATGCTC AAATAAAATC	4200
TAAATTGAC AATGTAAACC GAGTCGGATA GCTTTAAGTA CTGTTTGTAG GTTGAAGATA	4260
CGATTTTGA TAGGAACTCA TCAATTTTAG ATTTTAAAGC AGCATCAATA AATTGCTTCC	4320
TTGTTTGTG ATAATTTTTT TATTTAAAAA ATTATGACma GAGTGTGCTA TTCTTTTAT	4380
GAGAGGTGTA TGAATATGAT AAATGTATGT GATAAATGTA TGTGATGTTG GAAAAAGAAT	4440
AAAAGAACTT AGAATATCTT CAAATCTTAC TCAAGATAAG ATTGCTGAGT ATTTGTCTT	4500
GAATCAAAGC ATGATTGCCA AAATGAAAAA AGGTGAAAGG AATATCACGA ATGGATTTAA	4560
GTAATAAAGC TTCAAATCTT AGAAAAAAGT TGGGAGCTGA TGGTGAATCG CCGATAGATA	4620
TTTTTAAATT GCTACAAAAG ATAGAAAATT TGACGCTGGT ATTTTATGGA CTCGGAAAGA	4680
ATATTACCGG AGTCTGTTAT AAAGGAACTC AGTTCAGTCT CATTGCAGTC AATTCAGACA	4740
TGCCATTAGG AAGGTAAAGA TTTTCTTAG CACATGGACT GTATCATCTT TATTATGATG	4800
AGGTGAAGAA GAGTTCAGTC AGTCTTATCT TGATTGGTGA AGGAGATGAA ACTGAAAGAA	4860
AAGCGGATCA GTTGCTTCT TATTTTTTAA TTTTCCCATC TTCACTGTAT AGGATGGTTG	4920

401

AGGAAATCAG AGAAAATGCC AATAGAACTC ATCTTGAAGT AGAAGATATT ATAAAATTGG	4980
GTCAGTTTTA TGGTATCAGT CATAAAGCTA TGTTATATAG ATTGAGGAAT GATGGATACC	5040
TTGATGCAGA AGAAATTAAA AATATGGATA TTAGTGTTAT AGAGACAGCT TCAAGATTAG	5100
GCTATGATAC AAGTTTATAT CGTCCTTTGT CAGAAAGTAA AAAAGAAATG GCATTAGGAT	5160
AATATATTAA TTCAACTGAA CAACTTTTAG AAAATAACAG AATTTCCGAA GGGAAGTATG	5220
AGGAACTGTT ACTAGATGCT TTCAGATATG ATATTGTATA TGGGCTAGAT GAAGAGGGGG	5280
GAGTTGTCGT TTGACTAGTC GTGTATTTAT TGATGCAGAT TGTATTTTCTAG TATTTTTATG	5340
GGTTGGCACT GAACATCTTT TAGAAAAGCT CTATTTGGGT AAAATTGTTA TTCCACAAGA	5400
GGTGATGAT GAAATCAATA TACCTACAAT TCCCCATTTA AAATCTAGGA TAGATCAGTT	5460
GGTAGCTAAG GGTTCACTG AGATTGTGAG CATAGACATT GGAAGTGAAG AATACGCATT	5520
ATATAGAGAT TTAACAAGAA ATCATGATAG TAACAAGATT ATTGGTAAGG GAGAAGGGGC	5580
ATCTATTTCC TTAGCGAAAA AGCATAATGG GATATTAGGA AGTAATAACC TAAGAGATGT	5640
TAAATCATAT GTAGAAGAAT TTTCTTTAGA ATATATGACA ACAGGAGATA TACTGATTGA	5700
AGCGTTTAAA CCGTAATTTA TTAAGTGAATA AGAGGGCAAT CATATCTGGA ATAATATGCT	5760
TAAAAAGAGA AGGAAAATTG GTGCAAATTC ATTTTCAGAC TATCTTCGTG GAAGTATTCA	5820
TCAAAATAGA CAAAAATAAA TTTGGATAAA TCGAACTCAC TATTCAGGAG GCATATGAGC	5880
AATTCGAAAA AGAAAAGTGT CAAATTGAGC CTATAGGAGT AGAAGTGAAG TAGTAAGTCC	5940
TGCATAGTGG ATGAGAGAAA AGTTCTCCTT GAAGTTTCC TGAAGTATCA GTCGCATGTC	6000
AAACGATATG TAGGGTAATG TGAGAGGGGA TAGCGAGTAG TTTTGGTTA TTTTATCAAA	6060
AACTTATAT TTTATTATAC CGAATGATAA AATATATATA AAATGATAGA ATAAGGAAAA	6120
AACATGAATG TCAAAAAGAT AATGTCAATT TTTCAATCCT TTTATGTTGA TGTCAGTATT	6180
GAGGAACTGA CTTTGACTTT ACCAATCAGT TTTGTAAAAA GGTTTGAGTA TACTCAAATG	6240
ACTTTTCATA AGGAATCATT TTTATTGATT AAAGAAAAGA GAAGGGGGAG TTTGAGTTCA	6300
TTTGTTACTC AGGCTCGCAC TATGGGTGAA AAAGCCAATA TGGATGTTGT TTTGGTGTTC	6360
TGGAAGTTAT CAGACACTGA AAAAAAGCAA TTACTTCAAG CTAGAGTTCC GTTTGTAGAC	6420
TTTAAGGGAA ACCTCTTCTT CCCTCCATTG GGAAGTAGTAC TCAATGCGAA TGATACTGAA	6480
GTCCCTAAGG AATTAACACC TAGCGAACAA TTAACGTGGA TTGCCTTTTT ATTGACAAAA	6540
GGTCAAAAAG TAGTAGATGT TGATTGCTT TCACAAGTCA CTGGACTTCC AAATCAACA	6600
ATTTATAGGT GTTTGAGGAC TTTTAAAGCT TTATATGGT TAAACAAGCA AAATAAGCTT	6660



402

TACACATATA CCGTGTCAAA GAAAGAATTA TTCTTAAAT CCGTGTCAATG TTTATTTAAT	6720
CCCATCAAAA AACGGATTTT ATTGCCAGAT GCGATATAA AGCAGATAAA ATCTGTTTCT	6780
AACCTTCTAT ATGGTGGTGC TTATGCTTTG TCGCATTCAA CTTTTTTAGC TGAAACGGAT	6840
GAAAATATTA GCTATGTCAT ATGGCAGAGA AAATCAATC AGTTATCCTT GCCACTTTCT	6900
CAGCATGTTT TAAATGAAA GATGCTAGAG ATATGGAAAT ATCGTCCTTT TGTATCTGAG	6960
TTTTGGAATG ATTTTAAAA TAATCATGAT AAACAATTTG TAGATCCGAT TTCTCTTTAT	7020
TTGACCTTAA AAGATGATGA TGACCCACGT ATAGAGGAAG AGAGTGAAGC ACTAGAAAAT	7080
ATGATATTAC AGTATCTGGG AGAAGATGAT GCCAGCTAAT ACGAAAGTTA TTTTCAAGA	7140
AATGTTTGGC GATTTTCAGA ACTATTATGT TCTGATTGGG GGAAGTCTA CCTCTATCGT	7200
ATTGGATTCT CAAGGATTTA AAAGTCGCAC AACAAAGAT TATGATATGG TCATCATTGA	7260
TGAAGTAAAA AATAAGGAAT TTTATACTAC CTTGAATCAT TTTTAGAAT TGGGAGAGTA	7320
TCAAGGAAGT CAGAAAGATG AGAAAGCGCA GCTTTTTCGA TTTACAACAA CTAATCCTGA	7380
GTTCCTTCT ATGATTGAAC TATTTAGTAT CTTACCAGAA TATCCATTAA AGAAGGACGG	7440
TCGAGAAATT CCCTTACATT TTGACCAAGA TGCTAGTTTA TCAGCCTTAT TATTGGATGA	7500
AGATTATTAT AATATATTGG TGCATGAAAA AGAAACCATT CAGGGGTATT CGGTATTGAG	7560
TAATTGTGGT TTATACTCTT CGAAAATCTC TTCAAACCAC GTCAGCTTCC ATCTACAACC	7620
TCAAACAGT GTTTTGAGCA GCCTGCAGCT AGCTTCCTAG TTGCTCTTT GATTTTCATT	7680
GAGTATTAAT TATTTTAAAG GCTAAAGCTT GGCTGGATAT GAGGGAGCGC TCTGCCACAG	7740
GTGCTCAAGG TTTAAGTAAG TCCATTAAAA AGCATTTGAA TGACCTTACC CGTTTGACAG	7800
CTTCCTTGCT AGGAGATGAA AAGTTATCGG CTATAACATC AAGTAGTGGG GTAAAAGCAG	7860
ACATGCACCG CTTTGTGATA GAATTAGAGC CTGTGAAGTC AACTATTCTT CAAAATAATG	7920
ACATTTTATT GGATCAAAAT GAAATTTTGG AAATTTCTGAA AAATTTTCTC GATGGTTAAA	7980
ATAATTGTAG CGAGATGGCT ATATTGAATT CGTCTATATC TGGAACTAG AAAAACTTC	8040
AATTCAGGA GAAAATGAAG TCAATCTTCC CACAATCAAA CGTATAGTAT CAAGGTTTTT	8100
CAAGACCTGA TATTATGCGT TTTTGGCTTT TCAAACTTT TTGCCAGTC TTCGTTTTTA	8160
TCCTCTAGTC ACTTGATTG TTTCAGGTGG TTTTATAGTA TAGTAGAATG AAACGAGAAC	8220
AGGACAAATT GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAAGTGT	8280
ACTATTCTAG TTTCAATCTA CTATAGTTAA ATCTGCGCTC AAGTCTACTG GTGAATCTAT	8340
GATTGTAATA CTCTTCCAAA ATCTCATCAA CCACGTCAGT CTTGCCTTGC AGTCTGTATC	8400
TTACTGACCA AGCTAGTGAT GGATTTAGAA TAGGTGATTT GGAGCGTCTT ATTAGCTAGG	8460

403

AAATGCTGCT CATAGTCCTT TGCTGAGGCT AGGGTGTTTC AACATTCAAC ACTCAACTGG	8520
TTGATCTAGT TGATAGGAAG GGAGTTACTA TAAAATACTC AGGCTTCCAT CATATTTTTT	8580
GAAACGATTG TGTAATCAAA ATGTACCAAT ATTGTAGTAT TGGTACAGAA GATGTTGTGA	8640
ATGGATAAAT ATATCATAAC TGCTATCTCA AAAAGATTTC ATATGTCTGT GCATATATAA	8700
TAGACTTCCT GCAAAACTAG AATCCTAGTT CATGATTGAT AATACCAGCA ATCAAATTCA	8760
TTCGTAATCC AAAGCGTTTA CGATGATTTC GATAGGTTGT TGAAAACATT TTAAACGTTT	8820
CTACTTTGGC AAAGATGTTT TCAACCTTGC TTCTCTCCTT AGATAGCGCA TGGTTATAGG	8880
CTTTATCTTC AGCTGTTAGC GGCTTGAGTT TGCTGGATTT ACGTGGAGTT TGTGCTTGAG	8940
GACATATCTT CATGAGCCCT TGATAACCAC TGTGAGCCAA GATTTTACCA GCTTGTCCGA	9000
TATTTCTGCA ACTCATTTTG AACAACTTCA TATCATGACA ATAGTTCACA GTGATATCCA	9060
AAGAAACAAT TCTCCCTTGA CTTGTGACAA TCGCTTGAGC CTTCATAGCG TGAAATTTCT	9120
TTTTACCAGA ATCATTCGCT AATTCTTTTT TTAGGGCGAT TGATTTTTAC TTCCGTCGCA	9180
TCAATCATT CCGTGTCTTC AGAACTAAGA GGAGTTCTTG AAATCGTAAC ACCACTTTGA	9240
ACAAGAGTTA CTTCAACCCA TTGGCTCCGA CGGATTAAGT TGCTTTCGTG AATACCAAAA	9300
TCAGCCGCAA TTTCTTCATA AGTGCGGTAT TCTAGGCTTA ATTTAGGTTT TCGTCCACCT	9360
TTTGCGTGTT TAAGTTGATA AGCTGTTTTT AATACAGCTA ACATCTCTTT AAAAGTCGTG	9420
CGCTGAACAC CAACAAGACC CTTAAATCGT GTATCAGTTA ATTGTTTACT TGCTTCATAA	9480
TTTCGCAGGG AGTCTATTGA CTCTTTGGTA GGTGTCAATG TTTTTTTCAT CTATCCCGAG	9540
AATTATTTTC CCGCCATTTG TATTTGCAAA TCCTGAGTAG GTTCCCAGA AAGACTCTGG	9600
AAGATTGTTT TTAGCTTTTT TGTATTCTAA ATCAACCCCT TCAAATTTTA AGTCCATATT	9660
TTTCCTTTAC ATCTGTTTTT TGTGGTCTG GTATTTGTTT AAGTTGAGTG ATAATATAGC	9720
GAATTGAATT TCGAGAGTTT TTAATCAGTT AATTTCTTTT TTAACCCACT TTAATTGCTT	9780
TTTTAACACG GGTAAAAAA GAAATTAAAG TGGGTTAATT TTTCTTCA	9828

(2) INFORMATION FOR SEQ ID NO: 42:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3369 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

404

CCGCGAAAGA TATTTTGA CAAGAGTTTG GACGTGAGGT CCGGTTAT AATAAAGTAG	50
AAGTTGACGA GTTTTGTAGAC GATGTCATCA AGGACTATGA AACCTATGCT GCCTTGGTCA	120
AGTCACTTCG TCAGGAAATT GCGGATTTGA AGGAAGAATT AACTCGTAAA CCGAAACCTT	180
CACCAGTTCA AGCAGAACCC CTTGAAGCGG CAATTACAAG TTCTATGACG AATTTTGATA	240
TTTGTAAACG CCTGAATAGA TTGGAAAAAG AAGTTTTTGG TAAACAAAT TTAGATAACT	300
CAGATTTTGA AGTAGTTATT TGAGATGTGC AATTTTGGTA TAATCGCGTG AGGAGAATTG	360
TTTCTCATGA GGAAAGTCCA TGCTAGCACA GGCTGTGATG CCTGTAGTGT TTGTGCTAGG	420
CGAAACCATA AGCCTAGGGA CGAGAAATCG TTACGGCAGT TGAAATGGCT AAGTCCTTGG	480
ATAGGCCAGA GTAGGCTTGA AAGTGCCACA GTGACGGAGT CTTTCTGGAA ACAGAGAGAG	540
TGGAACGCGG TAAACCCCTC AAGCTAGCAA CCCAAATTTT GGTCCGGGCA TGGAGTACGC	600
GGAAACGAAC GTAGTATTCT GACTGCTATC AGCTAGAGCT GTTAGTGGTA GACAGATGAT	660
TATCGAAGGA AGTGGTCCTA GTCACCTCTG GAACAAAACA TGGCTTATAG AAAATTGCAT	720
ATAGGTTGGG GCTGAGAAAT TTTCTCAACC TCATTTTTTA AAGTGGACAT ATAGAAAGGT	780
CTTGCAAGAC TGTAACATGA AAAAAGAATT TAATTTAATT GCAACTGTGG CAGCAGGGCT	840
TGAGGCTGTC GTTGGTCGTG AAGTCCGAGA GTTGGGCTAC GATTGTCAGG TTGAAAATGG	900
ACGTGTTCTG TTTCAAGGAG ACGTGAGAGC TATTATCGAA ACCAACCTTT GGCTTCGGGC	960
AGCAGATCGT ATCAAAATTA TCGTAGGAAC GTTCCCAGCT AAGACTTTTG AAGAGCTATT	1020
TCAGGGAGTT TTCGCTTTGG ATTGGGAAAA TTATTTACCA CTTGGAGCTC GGTTCCCGAT	1080
TTCAAAAGCT AAATGTGTTA AGTCCAAACT TCACAATGAG CCCAGTGTTT AGGCTATTTT	1140
TAAGAAAGCT GTTGTCAAGA AATTGCAGAA AACTATGCT CGCCAGAAAG GGGTTCCTCT	1200
GATGGAGAAT GGCCAGAGT TTAAGATTGA GGTCTCTATT CTCAAAGATG TGGCAACTGT	1260
CATGATTGAT ACGACCGGT CTAGCCTCTT TAAACGTGGT TATCGTACCG AAAAAGGTGG	1320
CGCTCCTATC AAGGAAAATA TGGCAGCAGC CATTTTACAA CTTTCTAACT GGTATCCACA	1380
CAAGCCTTTG ATTGATCCGA CCTGTGGTTC GGGGACTTTC TGTATTGAGG CAGTTATGAT	1440
TGCTAGAAAG ATGGCGCCAG GTCTTCGTCT CTCTTTTGCA TTTGAGGAAT GGAAGTGGAT	1500
CAGCGATCGC TTGATTCAAG AAGTGCACAC AGAAGCGGCT AAAAAAGTAG ACCGTGAGCT	1560
TGAGCTGGAT ATCATGGGCT GTGATATTGA TGCTCGCATG GTGGAAATTG CTAAGGCCAA	1620
TGCTCAGGTA GCTGCTGTTG CAGGAGACAT TACTTTTAAG CAGATGCGCG TGCAGGATTT	1680
ACGTTCGGAT AAAATCAATG GAGTAATCAT TTCCAATCCG CCTTATGGTG AACGTTTGTC	1740
AGATGATGCA GGGGTGACCA AGCTCTATGC TGAGATGGGG CAAGTATTTG CACCGCTGAA	1800

405

AAC TTG GAGC AAATTTATCC TGA CTAGTGA TGAAGCTTTT GAAAGCAAGT ATGGTAGCCA	1860
AGCAGATAAG AAGCGTAAGT TATACAACGG AACCTTGAAA GTGGATCTAT ATCAATATTT	1920
TGGTCAGCGT GTCAAACGGC AAGAGGTAAA ATAGAAAGGG ATACTCATGA GTAAAAAAG	1980
ACGAAATCGT CATAAAAAAG AAGGTCAAGA ACCGCAATTT GATTTTGATG AAGCAAAAGA	2040
GCTAACAGTT GGTCAAGCTA TTCGTAAAAA TGAAGAAGTG GAATCAGGAG TCTTGCCTGA	2100
GGATTCCATT TTGGACAAGT ATGTTAAGCA ACACAGAGAT GAAATTGAGG CGGATAAGTT	2160
TGCGACTCGT CAATACAAAA AAGAGGAGTT CGTTGAACT CAGAGTCTGG ATGATTTAAT	2220
TCAAGAGATG CGTGAGGCTG TAGAGAAGTC AGAAGCTTCT TCGGAGGAAG TTCCATCTTC	2280
TGAAGACATC TTA CTACCTT TGCCTCTGGA CGATGAGGAG CAAGGCTTGG ATCCTCTATT	2340
GCTAGATGAT GAAAATCCAA CAGAAATGAC TGAAGAAGTG GAAGAGGAGC AAAACCTTTC	2400
TCGTCTGGAT CAAGAGGACT CAGAAAAGAA AAGTAAAAA GGCTTTATTT TGACCGTTTT	2460
GGCGCTTGTA TCAGTAATTA TTTGTGTCAG TGCTATTAT GTCTACCGTC AAGTGGCTCG	2520
TTCGACTAAG GAAATTGAAA CTTCTCAATC AACTACAGCC AATCAATCGG ATGTGGATGA	2580
TTTTAATACA CTTTATGACG CCTTTTACAC AGATAGCAAT AAAACGGCTT TGAAAAATAG	2640
CCAGTTTGAT AAAGTGAAGT AACTCAAGAC TTTACTTGAT AAGCTGGAAG GTAGTCGTGA	2700
ACATACGCTT GCCAAATCTA AATATGATAG TCTAGCAACG CAAATCAAGG CTATTCAAGA	2760
TGTCAATGCT CAATTTGAGA AACCAGCTAT TGTGGATGGT GTGTTGGATA CCAATGCCAA	2820
AGCCAAATCG GATGCTAAAT TTACGGATAT TAAACTGGA AATACGGAGC TTGATAAAGT	2880
GCTAGATAAG GCTATCAGTC TTGGTAAGAG CCAGCAAACA AGTACTTCTA GCTCAAGTTC	2940
AAGTCAAACT AGCAGCTCAA GTTCAAGTCA AGCAAGTTCA AATACGACTA GTGAGCCAAA	3000
ACCAAGTAGT TCAAATGAGA CTAGAAGTAG TCGCAGTGAA GTCAATATGG GTCTCTCGAG	3060
TGCAGGGGTT GCTGTTCAAA GAAGTGCCAG TCGTGTGGCC TATAATCAGT CTGCTATTGA	3120
TGATAGTAAT AACTCTGCCT GGGATTTTGC GGATGGTGTG TTGGAACAAA TTCTAGCGAC	3180
TTACAGTTCA CGTGGCTATA TCACTGGAGA CCAATATATC CTTGAACGTG TCAATATCGT	3240
TAACGGCAAT GGTATTACAA ACCTCTACAA GCCAGATGGA ACCTATCTCT TTACCCTTAA	3300
CTGTAAGACA GCCTACTTTG TCGGAAATGG CGCTGGTCAT GCGGATGACT TAGATTACTA	3360
AGCAGTCGG	3369

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9713 base pairs

406

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

AAGTTTACAA TTAAATGAA TTAACAATTT TCCCAACTAA AAGCACTCCA GTTACCGCAA	60
CGTTTGTA CT GAATGTACTA AATCGCATTC CATCAACTTC ATCTGTTTCG TCAACTTGAA	120
CAGATACTAA TTGAAGATTT AATACTTCTG CTGCCATAGC TAGCTCCTCC TATTTAAATT	180
TTTGGGATTA AGTACTTTAT CCACCCTCAT ATACTCTCTC CACCAGTAAA ATGCAAGCAA	240
TGATACAAAA TAGATTTAAC TATTTTATAT AGCGAAAACT TACAAATTTT TAAGAAATAA	300
TTTTTGCA TT CTAAAGATA AAATAGGAAC TTTTAGTAAT AAATATTAAA ATAAATAAAA	360
TAATAGATAC TATAAAATTT GGAAGTATTA ACCCCAAAAG ATTCATATCA TCTATTAAAA	420
TATCCTCTAA AGAGTAGTAT ATTAAAGCCA TAATTTTAAT GTTAAGTAAA AATGCAATTA	480
ATGAAGTAAC AAATGTCAA AATATAGCCT CACCAACTTT AATCTTAACC ATCTGGTAAT	540
TAGAAGTTCC TAAAAATTCA AATTGCTGAA TCTCAATCCT TTCTTGATGC GATGACAAAA	600
ATGCAATTGA AATAATATTT GCAAGTACTA TCAAAATTGG TGCTCCTACA TAGACAATAA	660
ATGCTACTTT TAGCTCTAAA TCACTGTCAT CTGAAATTG AGATAGTATA TTCTGAGAAA	720
TCATTTGAAA ACTAGAAATT AGTAATATAG CTCCTGTAAT TGCAGCACTG ATAGATTTTA	780
TATAAGACTT ACAATATAGT AAATTCCTCT TCGAAACAAT GAACATAAAA TTATTTCTAA	840
ATATAATTAT AGAAAGTAGT TTGATAAAAC ATGACTGTAT AAAAGGAGAT AATTGATAAA	900
TAATCACAAT ATCTAAGATT ACAATATTGA ATATTATCTG GGCCTTCGCT AAAATTGTGC	960
TATCTTGGAA AATTTGTTGC AAAGAAAGCA ACCAGATAAC ACTAAAACCA GCCAATAGCA	1020
GTATTCTTTT TACTATTGAA AGAACATGCC TTATTTTAGA ACTCTTCCTA TTTCTAATCT	1080
TCTTGAACGT ATAAAAGCAA CCACTTAGAA AGGCTAAAAA TGAAATCAAC ACTACTGTAA	1140
TGATACATCC AACAGCACTC GTTTGAAATT GGATATCAGG TAATATATTT TCCCCGAAAA	1200
AGTATTGTAA AAAATAATAA TAATTTGACG TAACAAATAT AGAGCATAGA TATGCAATAA	1260
AACTAATAAT CGAGGAAATG ATAAAAATCT GTCCCCCACC AAGAAATGAT AGTTGAAGGC	1320
GACTTGCTCC CAACACCTCC AGAAGTTCGT AATCATCTCT AAAAAATTCA ACCAACATAT	1380
TTATTATGTT AGAGAGCACA AAGAATAATG TTA CTCTCC GAATACTATC GGAAACATAA	1440
AAATGGGTTT AGGATCTGGA AGTCCGACAA ATACTTGCGA ATTATTCTCA ACATTAATTA	1500
CCCCATTAAC AGCCAATCCC ATA ACTAAAC TCGAAACAAA AATTACTGGT GAAACGCCTA	1560

407

ACCATTGTTT CTTATTATGT AAAAATTGAT AGTAACTAA TCTGAGCATC TCTATTCCTC	1620
CGTAGTTGAT TGTACCTCTA AGATTTTATA CAACTCTTCC CCGCTAGGTC TATGAAGTTC	1680
TTTGAAAATT TTTCCATCTT TCAATATTAA TGCACGATCA GTTTTCGAGG CCAATTCTAT	1740
ATCGTGC GTT ACCATAATTA CACACTTACC CGCCCTACT AACTCTCTCA ATAATTCAAA	1800
AATTACTTCA CGAGAAACGC TGTCTAAAGC CCCAGTTGGC TCATCAGCAA ATATTATATC	1860
ACTATCAGCA ATAACCGCTC TAGCTATAGC AACCTTCTGT TGTCTCCAC CAGACAGAGT	1920
TCCAACAAAA TCGTTTAAGC CAGCATTAAT CTTCATTCTT TTGAGTAAGT TTTCTACATT	1980
TTAATAGTT AATTTTTTTT GTGATAATCG CAAAGGAAGT GCTATATTTT CTATTACCGG	2040
CAGGGAAGGT ATTAAATTGT ATGCTTGAAT TATAAAGAT ACTTCGTTAC GTCTTATACT	2100
TGACAATTTT GCATTCTCTGA TTTTATAGGG GTTGATTCCA TTTAAATTA CTTCCTCACT	2160
TGTTGGTCA AGCAAACCTAG AAATACATTT TAATAAGTT GACTTTCAG AACCACTAAT	2220
TCCTAGAATA CTTATAAATT CTCCTCTCGA AGCAGAAAGA GAAACATTTT TCAGCACTTG	2280
CAACGTTTTA TTATTCCTA GTAAAAATTG ATGATACAGC CCTTTCACCT TTAATATATA	2340
ATCTTTATCC ATATTCTTGC CTCCAATCAC TTAATTTTGA AAAGTGTCC ATTTTCCAAT	2400
TTATATATAT CAGTGTATCT CTTGTCATTT AAGTCATAAT GATGTGAAAC TTCAATAAAT	2460
GAAATACCTA AATTGAACAG AATATCATGT ATGGAATTTG AATTATCATT ATCTAAATTA	2520
GCTGATATTT CGTCAAATAA GTACACTTTA TTATTTCTAA TCAGACCTCT AGCTAAAGCT	2580
ATTTTGTGTT TTTGACCTCC AGACAAATTA CTACCATTTT CACCACATTG ATAATTTAGT	2640
ATATCTATCT TTTCTAATTC TTCATATAGA TTTACCTTTT TTAACACCTC AATTATCTGA	2700
TCATCTGAAA AATATTCATT TTGAAATAAA GTTACGTTCT CACGAATAGT AGTGTCAAAA	2760
ATATATGGTG TCTGATCAAC TGTGGTATT GAATCTGAAC TCTTTTCCC ATGTGATAAC	2820
AAATTTACAT AACCTTTTGG TGGCTTTAAA GAACCATTA TTAATTTAA AATCGTTGTT	2880
TTCCCACTAC CAGAAGTTC TGTTAATAAT ACCCTAAATG GTGACTTAAA TGAGAAGTCA	2940
ATACTTAATT TATTTTCTGG TGTAATAGAA TATACAACAT CTTTCATGTG TATCTCATCT	3000
ATTGATGAAG TATACAGTCC GTTATTATCA TGTTCAGCGT CTATAAATTT CTTCTCTCCA	3060
CTTAAGTATT TTA AAAACGG TTTCTTAAA TCTTTGGTTG TATTTATCTT ATTTAATGAA	3120
TAGGCAATTG ATTGTATCGG CCCTAAACT TTATCGTTTG CTAAGAAAAT ACCTATCAGT	3180
TCACATAAAG AAAGGCTTTT ATGATAAATT ACAAATAAC ATCTACAAC CAAGGGAAGT	3240
AGAAAGCAAA AACCTGAAAT TAGTACTGCA ACCAATTTTG AAAGAACCTC TGATCGTTTC	3300

408

AAATTAAAAG TAGAATCTTC TAGTTTATCC AACTTTTTTAT CCGACAAACT AATTATTTCT	3360
TTAGTAACAG AATAAGATTT TAATGTCTTA AAACCATTAA AAATTTCTTT TATTATGTGA	3420
GTATACTCTG CATTGCTGTT AGAGTACTCA TTAGCTGAAT TAGACAACAT CTTCTTCATA	3480
AAGACAGGTA CTATAATCGG CAATGCTGAT AATACAATAA ATATTATTGA AACTAGGAAG	3540
TTTAAATAAA GCATAAAACT TAGAGAGACG ATGAACAACA ATATTGAAGA AATTATTTCA	3600
AAAATTTGTC TAAAATAGTT TTCTTCGATT AATCTCAAAT CATTTGACAA AACTGAAATA	3660
ATAGATGAGT AATCTTTAAC CATTTGAGAA GAAAGATACT GTTCTCTAAA ATATCCTTGT	3720
TTAATTTTTA CATTTATATC TTTAGTTATT GATGCTTCCG TTACTTCTAA ATAGTAATTT	3780
GATATATAGA TTGCTGACCA ACCCAGAATA CTTATAGCAC CAAATCTTAG AACGTCAGAA	3840
AATGAGGAAG TCTGATTTAA ACTACCTGCA TATACAATAA TTCCTGAGAG CAAGACACCA	3900
TTAAACGAAG ATAGAAATAT TAAATCCCC ATTAATATAA GTTTAGTCTT TTTTATAAAT	3960
TTTAAATAAT TCATAAGTTA TTCCTTCCCA CTTCTTCAAA GAAATAATTT AAAGTATCAA	4020
TCATTAAGAG AACATCTGAT GGAGTAAAC CTCCATGACC AGCTGCTTTG TTTAAATACA	4080
ACAAACTTTT AACTCCAATA GAATTTAATT TCTTTGACCA CTCTATCACT TCGTTATTAT	4140
TAATATATGG GTCTTTCTCA CCCAAAATAT TAACTATAAC AGTATTTGAG TCTCGTGCCT	4200
TTTCAATATT TTGCATAGGC GAATATGACT TTATATAAGC CTTTACTTCA GGGTCTCTAA	4260
TATCTCCCCA CTCTGCTATT TCGGTCTTAG AAAGAGGATC ATTTGGATTG TGAAGTGTAT	4320
CATAAGGATT TATAAATGGC GAAAATAAGA GAATGCTTTG CAATAAATTT TTTTCTCGT	4380
TCAACACCGC ACCAGCAATT ATTCCACCTG CACTAGAAGT TATTAAACCT AATCGCTTAC	4440
TGTCAATTAC ATCATTTTCC CTTAAATAAT TTAATCCCTC AATAAAATCT CTGATAGAAT	4500
TCCAITTGTT TAACGCCTTT CCTGAGCGAT ACCATTCAAC ACCCAAATAG CCTCCACCTC	4560
TTACATGAAC TATAGCATAA ATAAAACCTG CATCTATTAT AGATAACATA ATTTCATCTA	4620
AATCAGAATT ATCATTTCTTA CCATAAGCCC CATAGACACT TAGAATACAT TTTTTCTTTC	4680
TTGGGAGCTC ATCCGTATCT TCACTTTTCC AAAATAAAGA AATCGGTATG CTTACATCAT	4740
AACTGTCTTT TTTAGTCCAA ATCACCTTAG AAAAATATTT AGTATTATTC GATTTTATGA	4800
TGGGTCTTTC AAATTCAGTT TTTAATGTAT TTTCTATTAA ATCAAAACTA AGTATTTTTT	4860
CGTAAAAAGT TCTCCTCTCT AAAAACAGAA GAACACGATC AGAAAAATGAA TTTTCATAAA	4920
GTGTTGTCTT TTCATCAAAT GTTATCTTAT TAACACTCAA CTCCCTCAAA CTATTATTTT	4980
TAAATGTAGC AAGATAAAAG ACGGAATTCT CTGCGTTTGA ACAGTCTAAA AGGATATAAC	5040
GTCCTATACA GTGAACTCTT CTAGCCCTAT CTTGATATGG TATAGTAATA GAAACTCTGT	5100

CTCCCGAAGA AGTTTCCCTT AGAATTAGTT GATCTTCTT TTCTTCAGTT GAAGAGAGCC 5160  
CAAGAAAGTA CTGTGCTTTT TCTGTACTAA ATAGAGCGAT ATCTCTAGGT GTTGGGGCTA 5220  
CCGTTTCTGT GTAAGAGTGT CTAACAAAAC CCGTCCGGTC GAAACTGTAT AGAAAAATCC 5280  
TGCCTTCTG AAAGTCTACT GACTTTACAA AACAATTATT GCTATCAATG TGGACTATTT 5340  
TTAATCGAAA AGAGCATTCG TTTTCTTCAA ACAGTTCCTC TTCTGTAAAG CTATCAAAAG 5400  
ATTTATAGAA TAACTTACTT GGCCTCCCGT ACTCTTTGGA GCGAGTATAC ATAACACCGA 5460  
ATTTACCCAA ATAGAACGAA CTTTCTACTG AAATATCTTC AATGATAAAT AACTCTTCCA 5520  
TAGTATATTT TTTTATTCCA ATTAAATTAG TCGTACGCAG TGAGGATACA ACCAAAACTA 5580  
TATAACTCTC ATCAGATGAA ATCCTAACAT CCTGTAAGAT ACTATCATCT GGCAAAAGTAT 5640  
ATTTTCCAC ATCAAAGACA ATTTAAGTG AATTGAATT GTCTAACTG GAAGAACTAA 5700  
CCTTAGGAAT CCAGTCATTA TCTTCGACAT ACCATTCCTT TATTACACCA GTATTGGGTA 5760  
TACTCCAATT ATCAAATTGG TACCAATATC GCCCTCTCCT AAATATCAAA GAATTCCATT 5820  
TTTTTAATTC CTGAAATGAT GAAGAGATAG ACCTCTTATA GTGTGTTTTT TCCTGTATTG 5880  
TATTTAAAAA TATTTTATTA CTCTGATTCA CAAGTATGAC CCCTTAATAA TGGTATCTAA 5940  
ATATTATATT TGAGGAAGAA TCGTCAATTT ATTATCCATT ATTGATACCA ATCCAATTGC 6000  
AACACCCGCA AATCCCGAAG CAATATCTGT TGTATCTTT AAACCATTAT CTCCCGCAAT 6060  
AACAAATCCT TCTTCAATTA CACACAAATA TCTATAAAGT TGTTCATTA ATTTCTTTTG 6120  
TCCTGAAAAG TTATCATCGA TATCACTATA TATATTATTA GCAACTTCAA GACCACAAAA 6180  
TCCGTAAAT AAACCTGGTA ATACACAAAA AACTACATCA GTTGCCCTCT CTAAGAAGT 6240  
TAAATATTTT AAGTATTTGC TTGACAAGAT TTCTTTATTT CTATTAATAA GTAAAAGCAG 6300  
GCCAGCACTT CCAGTTGCTA GATATGGTAG TAATCTATGA CCTTGGCTGT ACTGCAATGA 6360  
ATTATTACTA TCTACTTTAT AAGCAACTAA TTCTTTATCT ACAGCCAATT CTAGACCATT 6420  
TTTATAGATA CTTTCACCAG TTAATTTATA AGCTTCACCG AAGAGCCAAG CTACCCCTGC 6480  
GTGACCATAT AGTAATCCAC CAAAATTCTC ATAAGGATCG TTAATCTGAA CATCACTAGC 6540  
GCCAACTTTA CAAAAAGTTT CTGGATTTTC TATATAATTT AAAGTATATT CTCTAAGCCT 6600  
AATTAGTATT TCTTCTCCTA GTTTATTATC AATTCCCTCT TTAATAAGAA AATACAGTCC 6660  
AACCAGTAAA ATTCCAGCCT GCCCACTATA TAAATTTTAA TTTTGTGAAT TCTCAAATAT 6720  
CTCTATAAAA TGAGTTGTAA AAAGTTCAAC TGCCCGATCT ATCTCCCAA ATTCAATAAT 6780  
GAGCCAGATT GTACCAATTT TACCATCAAA AAGACCAGAA AGGGACGATT TCTTAAATTT 5840



410

ATTTACTGCC TCATTAATAA CCTGTGTTTCG AATCTCATAA TAGTCATCAA ACTTGAAATT	6900
TTTTACTTTC TTAGCTAGTT GTTGATAACT CCAAAGGATA GCTAAATCTG AAAACGCAAT	6960
TCCTTGATTA AAATTCAGAC CATAATAATG AACTGGGAAG AATCTTGATT GAAATTCTTT	7020
ACGCCACTGT CCATAAGTTA GCGTAAACCC TCTCAATAAT TTTATAATAA AATCTTGAT	7080
ATCTTGCTCA CTCTCGATAG TTCTAATCTC ATGCATGGGT TTTAAAACCT TTTCTCGGA	7140
AATATTCTCA ATCTGTGGAC ATTTAGAATC TAGATATGAC AATAAACTTT CTACATAATC	7200
TATATGTTCT CTTGTATAAC CCAAAGACTC AAATAGTTTT TTTCTTCTA TCCTGGTTTG	7260
ACTTACATAG TTGTATGTCA AATCCGATGT AGTTACTAGT GGCATGTATA AATAATGAGC	7320
TATTTGTCTA ATACCATAACC AATCTATCTC ACTGGGAAGT GTTCTCGCC ATGCTCTAAA	7380
ACCAGGGGCT GCAACTTTAT GTACAACCTT TTCATCATTT GAAAAGACAG CCTGTTCCCA	7440
GTCTATTATA CTAATCTCAT CTTTCATCCTT AACCAAGATA TTTCTAAAT GTAAATCTTG	7500
ATGATATACA TTTTCAGAAT GAAACTTATT CGTTAAATCG ATGAGTTTTT CTACTATCTT	7560
TGAAACTCTC AATAGATAAT CTTTGGTCTT ATCAACAACT TCATATAAAG GAAAATTATT	7620
GGTAACCCAT CTATTTAGTG GAACGCCCTT CATATGTTCA ATTCCTAAGA AGGTGTGCTC	7680
CCAGATCTTA CCGTGCCAGT ATATTTTAGG CGTCTCACTC CATTCAATTA GAATTTTATAG	7740
TGCTTTGCAC TCCGAAGCTA ATTTCTCTGA AGAATAAGTA CCATCAAATC CTAGACCTGT	7800
ATACGGTCTA GCCTCTTTTA AAATTATTTT TTTCCCATCT TCTTTTAGCC TAGCATTATA	7860
TATCCCACCA CTGTTTGAAA ATCTAATTGC ATTATCTATA ATAAAGGGAA AGTCTCCCTG	7920
TTTTTTATCT TTCTTGTCAA GCCATTTATT CAAAAAGTCA GGGGGCACTA TACCTTTTGG	7980
AATTTTAAAT ACTGGTAAAC GTTCATCTTT AACAACCTCA TCGCCAACAA TTAATTCATC	9040
AATAGCAACC TTCTTTTCAT CATCCCTTGA CGGCCTAAAC ACACCATACC TCAGATATAT	8100
TGGTGCTTCA TCCCAACGTT TATCGCTTAA AATATATGGC CCATTATATT GCTTTAAGGC	8160
ACTTTCTAAC CTTTGCAAAA CCGACTCTAA TTCATTTTGA TTTGGATAAC ATGTAATAAA	8220
TTTACCAGAA AATCCTCGAC TAACCAATTT CCCGTTTCGC ATGATAAATT TGTCTTCTGT	8280
ACTAAGATGT TTAAATGGAA TTCGCATTTT ATGGCAAATT TTTGCTACAT CTTGTAACAA	8340
TTCATGTGAA CTGTTATACT CTGAACATAA GTGTATTTTC CACCCTTGTC TTTCAACAAA	8400
TTTTCCAATA GGGTATTGAT AAACCCACTC ATCATTATTC ATTACTTCGT GCCAATTAAT	8460
AGGCAGACTT ACTTGGTACT TTATGCTAGT ATCTGCTACTA TAATCATTAT TAGTGAAAAA	8520
GAAAGGATGC TCCAAATTGA AATTATAATC CATAACAAAA TCTCCAAGAA ATTTTATCAA	8580
ACTTAATATA TCTATAGCTA GACAGACTTA TTTAAATAAA AAGGGAGAA CTTTTGGATT	8640

411

CTCCCATAT AAGCACTAAC ATTCCAACGT GCACATATTG GAACGACATC CATAACTCCA 8700  
 GAGAACTCTT AAAGTTTACA ATTTAAATGA ATTAACAATT TTCCCAACTA AAAGCACTCC 8760  
 AGTTACCGCA ACGATTGTGA CTGAATGTAC TAAATCGCAT TCCATCAACT TCATCTGTTT 8820  
 CGTCAACTTG AACAGATACT AATTGAAGAT TTAATACTTC TTCTGCCATA GCTAGCTCCT 8880  
 CCTATTAAAA TTTTGGGAT TAAGTACTTT ATCCACCCTC ATTATACTCT CTCCACCAGT 8940  
 AAAATGCAAG CAATTATACA ATGTTGTCAC ATAGAAAATA ATGTTTCCGT AACTTTTCAA 9000  
 AGTAACTTCC ATCTCTCTCC CAAAAGTGA AGTTAGTTTT AGAAGTTACC TAAAAATCAG 9060  
 GTCACCTATT TAAAAAAGC AGCAAACTAT AACTAGTAG GTTCCACACC AAATGTAGTC 9120  
 CCATACTGCC CCATAAGTCA GATTTATAGC GCACCATACC TAAAAACATC CCAAGTGAAA 9180  
 CATACAAACA CCAAGCTAGA ATGGTTCCTG TATGATGTGC TAAGGCAAAT AAAACACTTG 9240  
 TCAAAGCAAC TCTGATATCT AATTTTCTGA CCAAATTCCA TAAAATTCTT CGATACAGAA 9300  
 ATTCTTCAAC CATACTCGCA TTGATTAAGA ACAATAAAAA TGAAAACCAA GGAATTGAT 9360  
 GTTGAAGGCC AATTAAGTTT GCTTGATTCTG TGCTTCCTTG AGCATGAATC AGACTAAAAC 9420  
 ATAGACTTAT AATCAGTAGG CTAACAAATT CAACACCAAG CCATTTTCATC CTAGATTTCA 9480  
 TATTGACCTT ATGCGCTTGT TTGCGTTGGC CATACATCCA TAAAAAGAA ATGAGTGACG 9540  
 AACCATAGAG AATCTGTAGT ATAGTTTACT CACCGATACA AAGAAATTTC AATAAGTATA 9600  
 GAGrTACCAA TAsGACATTT ACTTGTTGGA ATATATAAAC TGGAATTATT CTTTTCATAG 9660  
 TTACCTCCGA AATAAATCTT CATAATCTAA ATCTAATACC TGCACAATCC TTT 9713

(2) INFORMATION FOR SEQ ID NO: 44:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 8657 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

AAAGAAATTG TCAGAGAGTG GCTAGATGAA GTAGCAGAGC GGGCTAAGGA CTATCCAGAG 60  
 TGGGTGGATG TTTTCGAGCG TTGCTACACC GATACCTTGG ACAATACGGT TGAAATCTTA 120  
 GAAGATGGTT CAACTTTTGT CTTGACTGGG GATATTCCTG CCATGTGGCT TCGAGATTCT 180  
 ACAGCCCAAC TCAGACCCTA CCTTCATGTA GCTAAAAGAG ATGCCCTCCT GCGTCAGACC 240  
 ATTGCAGGTT TGGTCAAACG TCAGATGACC TTGGTACTCA AGGATCCCTA TGCTAACTCC 300

412

TTCAACATTG AGGAGAACTG GAAAGGGCAC CACGAGACTG ACCACACAGA CCTTAACGGC	360
TGGATCTGGG AGCGCAAGTA TGAGGTGGAT TCGCTTTGCT ATCCTTTGCA GTTGGCTTAT	420
CTCCTCTGGA AAGAGACTGG CGAGACTAGT CAGTTTGATG AGATTTTTGT CGCAGCGACT	480
AAGGAAATTC TCCATCTGTG GACGGTGGA CAAGACCACA AGAACTCTCC TTATCGTTTT	540
GTCCGAGATA CGGACCGTAA GGAAGACACC TTGGTAAATG ATGGCTTTGG ACCTGACTTT	600
GCAGTGACAG GTATGACTTG GTCAGCTTTT CGTCCGAGTG ATGACTGTTG CCAGTATAGT	660
TACTTGATTC CGTCAAATAT GTTTGCTGTA GTAGTCTTGG GTTATGTGCA AGAAATCTTC	720
GCAGCATTA ACCTAGCTGA TAGCCAGAGT GTTATTGCTG ATGCCAAGCG TCTTCAGGAT	780
GAAATCCAAG AAGGAATCAA AACTACGCT TACACCACCA ACAGCAAGGG CGAAAAGATT	840
TACGCTTTTG AAGTGGATGG CCTAGGAAAT GCCAGCATCA TGGATGATCC AAATGTACCA	900
AGTCTACTAG CTGCGCCCTA TCTGGGCTAC TGTTCGGTCG ATGATGAAGT GTATCAAGCT	960
ACTCGTCGTA CCATTTTGAG CTCTGAAAT CCATACTTCT ACCAAGGAGA ATACGCAAGC	1020
GGTCTCGGCA GTTCTCATAC CTCTATCGC TATATCTGGC CAATCGCCCT TTCTATCCAA	1080
GGCTTGACAA CAAGAGATAA GGCAGAGAAA AAATTCTTGC TGGATCAGCT GGTTCCTTGC	1140
GATGGTGGTA CAGGTGTCAT GCACGAAAGC TTTCATGTAG ATGATCCGAC CCTCTACTCT	1200
CGTGAATGGT TCTCCTGGGC TAACATGATG TTCTGTGAGT TGGTCTTGGA TTACTTGGAT	1260
ATTGCTAAG GGGCTCGCTT TAGCTCAACC GATTCTTATC AGAATCACAA GTTTACATTT	1320
AAAACGTTAA AATTTAAATT TAGAATGAGG TTTTACTTCA TGGAAAATGT TGTTGTACAT	1380
ATTATCTCAC ATAGTCACTG GGATCGTGAG TGGTACTTGC CTTTTGAAAG CCATCGTATG	1440
CAGTTGGTGG AATTGTTTGA CAATCTCTTT GATCTCTTTG AAAATGACCC TGAGTTCAAG	1500
AGTTTCCACT TGGATGGACA AACTATTGTC CTGTATGACT ACTTACAAAT TCGCCCTGAA	1560
AATCGCGACA AGGTCCAACG CTACATTGAC GAGGGCAAAC TTAATAATTGG TCCCTTTTAC	1620
ATCTTGCAGG ATGACTACTT GATCTCCAGT GAAGCCAATG TCCGCAATAC CTTGATTGGT	1680
CAACAAGAAG CTGCCAAATG GGGTAAATCA ACCCAGATTG GCTACTTTCC AGATACCTTT	1740
GGAAATATGG GACAAGCGCC TCAAATTCTT CAAAAATCAG GCATTACGT GCGGGCCTTT	1800
GGTCGTGGTG TGAAGCCGAT TGGATTGAC AACCAAGTCC TTGAAGATGA GCAGTTTACG	1860
TCTCAGTTTT CAGAAATGTA CTGGCAGGGT GTGGATGGTA GTCGTGTTTT AGGTATTCTC	1920
TTTGCCAACT GGTACAGTAA CGGGAATGAA ATTCCAGTTG ACAAAGATGA GGCCTTGACC	1980
TTCTGGAAAC AAAAAATTGTC AGATGTGCGT GCCTACGCTT CGACCAACCA ATGGTTGATG	2040
ATGAACGGCT GTGACCACCA GCCTGTACAG AAAAAATCTGA GCGAAGCCAT TCGTGTGGCA	2100

413

AATGAACTCT TCCCGGATGT AATCTTTGTT CATAGTTCTT TTGATGAATA TGTTCAGCT	2160
GTAGAAGGTG CGCTTCCTGA ACACTTATCA ACTGTTACAG GCGAGTTGAC CAGTCAGGAA	2220
ACAGATGGCT GGTACACACT TGCCAACACT TCTTCATCCC GCATTTACCT AAAACAAGCC	2280
TTCCAAGAAA ATAGCAACCT CCTAGAGCAA GTGGTAGAAC CCTTGACTAT TATCACTGGT	2340
GGACACAACC ACAAGGACCA GTTGACCTAT GCTTGAAAA CACTTTTGCA GAATGCGCCA	2400
CATGATAGTA TCTGTGGCTG TAGCGTGGAC GAAGTTCACC GCGAGATGGA AACGCGTTT	2460
GCCAAGGTCA ACCAAGTAGG AAACCTTTGTT AAAAGTAACT TGCTCAACGA GTGGAAGGGT	2520
AAAATTGCTA CGGATAAGGC TCAAAGTGAC TATCTCTTTA CTGTCAATTA CACAGGCTTG	2580
CATGATAAGG TCGATACTGT CAGCACAGTG ATTGATGTGG CGACTTGTGA TTTCAAGGAA	2640
TTGCACCCAA CAGAAGGCTA CAAAAAGATG GCTGCTCTTA TCTTGCCAAG TTACCGTGTG	2700
GAGGACTTGG ATGGTCGTCC TGTAGAGGCT ACAATCGAAG ACCTCGGAGC TAATTTTGAG	2760
TATAATTTAC CAAAAGACAA GTTCCGCCAA GCTCGTATTG CTCGTCAAGT GCGCGTGACC	2820
ATTCCAGTTC ACCTAGCGCC GCTTCTCTGG ACAACCTTCC AATTGCTGGA AGGAAAACAA	2880
GAACACCGTG AGGGTATTTA CAAAACGGA GTGATTGATA CACCATTCTG AACGGTGAGT	2940
GTGGATGACA ACATCACAGT CTATGACAAG ACAACTCAGC AAGCCTATGA AGACTTTATC	3000
CGCTTTGAAG ACCGTGGGGA CATCGGAAAC GAGTATATCT ATTTCCAACC AAAAGGAACA	3060
GAGCCAACTT TTGCAGAGCT TAAGGGCCAC GAGGTCTTGG AAAACACAGC TTGCTATGCT	3120
AAAATCTTGC TCAAACATGA ATTGACCGTG CCTGTCACTG CGGATGAAAA GCTAGAAGAA	3180
GAGCAACAAG GTATCATCGA GTTTATGAAG CGTGAGGCTG GACGGTCAGA AGAATTGACA	3240
AACATTCCCTC TGAAAACCTGA GTTGACTGTC TTCGTTGACA ATCCACAAAT CCGCTTCAAG	3300
ACTCGCTTTA CTAACACTGC CAAGGATCAC CGTATCCGTC TCTTGGTCAA GACTCATAAC	3360
ACGCGTCCAA GCAATGATTC TGAAAGTATC TATGAGGTGG TGACACGACC AAACAAACCA	3420
GCTGCTTCAT GGGAAAACCC TGAAAATCCT CAACACCAAC AAGCTTTTCT CAGTCTGTAT	3480
GACGATGAAA AAGGGGTGAC TGTATCCAAC AAGGGATTGA ATGAATACGA AATCCTTGGG	3540
GATAACACCA TTGCCGTGAC CATTTTCCGT GCATCAGGTG AGCTAGGTGA CTGGGGCTAC	3600
TTCCCAACGC CAGAAGCACA ATGCTTGCGG GAGTTTGAAG TCGAGTTTGC ACTTGAATGC	3660
CACCAAGCCC AAGAACGCTT CTCAGCTTAT CGTCGTGCCA AAGCCTTGCA GACACCGTTT	3720
ACCAGCCTTC AGCTTGCTAG ACAGGAAGGA AGCGTGGTTC CGACTGGTAG CCTCTTGAGC	3780
CATTCTGTTT TCAGCATACC GCAAGTTTGT CCAACAGCCT TTAAGGTAGC TGAAAATGAA	3840

414

GAAGGCTATG TGCTTCGTTA CTACAATATG TGTAGTGAAA ATGTACGTGT GCCAGAAAAGT 3900  
CAACATCTCT TCCTTGACCT ACTTGAACGA CCATACCCAG TTCATTTCAGG ACTATTGGCT 3960  
CCACAAGAGA TTCGTACAGA ATTCATCAAA AAAGAAGAAA TTAAATTTCA AAAAGTAAAC 4020  
ATCAAAAGAA AGGAGGGGCG AAAAAGTAAG AACTAACTGC TGATTGCCCC CTTTATGGT 4080  
AAAAACAATG ACCATTGCAA CGATTGATAT CGGAGGGACT GGGATTAAAT TTGCCAGTCT 4140  
GACTCCTGAT GGGAAAATAC TGGATAAGAC AAGTATTTCA ACGCCTGAAA ACTTGGAGGA 4200  
TTTACTAGCG TGGCTAGATC AACGCTTGTC AGAACAGGAT TACAGTGGGA TTGCTATGAG 4260  
CGTTCCAGGT GCAGTCAATC AAGAGACAGG TGTGATTGAT GGCTTCAGTG CGGTGCCCTA 4320  
CATCCATGGC TTTTCTTGGT ATGAGGCCCT TAGCTCTTAT CAGCTACCTG TCCATTTAGA 4380  
AAATGATGCC AACTGCGTTG GACTCAGTGA ACTACTAGCT CATCCAGAGC TTGAAAATGC 4440  
AGCCTGTGTC GTGATTGGGA CAGGGATTGG CGGAGCCATG ATTATCAATG GTAGACTTCA 4500  
TCGAGGTCGC CACGGTCTGG GTGGAGAATT TGGCTACATG ACAACCCTTG CCCCTGCTGA 4560  
AAAACCTAAT AACTGGTCGC AACTAGCATC AACTGGGAAT ATGGTACGAT ACGTGATTGA 4620  
AAAATCTGGT CATACTGATT GGGACGGTCG CAAGATTAC CAAGAGGCCG CAGCTGGTAA 4680  
TATCCTTTGT CAAGAAGCCA TTGAGCGCAT GAACCGCAAT CTGGCGCAAG GCTTGCTCAA 4740  
TATCCAGTAT CTGATCGATC CAGGTGTCAT CAGTCTGGT GGCTCTATCA GTCAAAATCC 4800  
AGATTTTATC CAAGGTGTCA AGAAGGCTGT TGAAGACTTT GTCGATGCCT ACGAAGAATA 4860  
CACGGTCGCA CCAGTTATCC AGGCCTGCAC CTATCACGCA GATGCCAATC TCTACGGTGC 4920  
TCTGTCAAC TGGCTACAGG AGGAAAAGCA ATGGTAAGAT TTACAGGACT TAGTCTCAAA 4980  
CAAACGCAAG CTATTGAGGT TTTAAAAGGT CACATTTCTC TACCAGATGT GGAAGTGGCT 5040  
GTCACCTCAGT CTGACCAAGC ATCTATCTCT ATCGAGGGTG AGGAAGGTCA CTATCAATTG 5100  
ACCTACCGCA AACCTCACCA ACTTTATCGT GCCTTGTCCT TGTGGTAAC AGTTCTAGCA 5160  
GAAGCTGATA AAGTAGAGAT TGAGGAACAA GCAGCTTACG AAGATTTGGC TTACATGGTT 5220  
GACTGTTCTC GAAATGCGGT GCTGAATGTG GCTTCTGCCA AGCAGATGAT TGAGATATTG 5280  
GCTCTCATGG GCTACTCAAC CTTTGAGCTT TACATGGAAG ACACTTACCA GATTGAAGGG 5340  
CAGCCTTACT TTGGCTATTT CCGTGGAGCT TATTTCAGCAG AGGAGTTGCA GGAATCGAA 5400  
GCCTATGCCC AACAGTTTGA CGTGACCTTT GTACCATGCA TCCAGACCTT GGGCCACTTG 5460  
TCGGCCTTTG TCAAATGGGG TGTCAAGGAA GTGCAGGAGC TCCGTGATGT AGAGGACATT 5520  
CTTCTCATTG GCGAAGAAAA GGTATATGAC TTGATTGATG GCATGTTTGC CACGTGTCTT 5580  
AAACTGAAGA CTCGCAAGGT CAATATCGGG ATGGACGAAG CCCACTTGGT TGGTTGGGA 5640

415

CGCTACCTGA TTCTGAACGG TGTGTGGAT CGTAGTCTCC TCATGTGCCA ACACCTGGAG	5700
CGCGTGCTGG ATATTGCTGA CAAATATGGT TTCCACTGCC AGATGTGGAG TGATATGTTT	5760
TTCAAACCTCA TGTCAGCGGA TGGCCAGTAC GACCGTGATG TGGAAATTCC AGAGGAAACT	5820
CGTGTCTACC TAGACCGTCT CAAAGACCGT GTGACTCTGG TTTACTGGGA TTATTATCAG	5880
GATAGCGAGG AAAAATACAA CCGTAATTC CGCAATCATC ACAAGATTAG CCATGACCTT	5940
GCATTTGCAG GGGGAGCTTG GAAGTGGATT GGCTTTACAC CTCACAACCA TTTTAGCCGT	6000
CTAGTGGCTA TCGAGGCTAA TAAAGCCTGC CGTGCCAAATC AGATTAAAGA ACTCATCGTA	6060
ACGGGTTGGG GAGACAATGG TGGTGAACT GCCCAGTTCT CTATCCTACC AAGCTTGCAA	6120
ATCTGGGCAG AACTCAGCTA TCGCAATGAC CTAGATGGTT TGTCTGCGCA CTTCAAGACC	6180
AATACTGGTC TAACGGTTGA GGATTTTATG CAGATTGACC TTGCCAACCT CTTACCAGAC	6240
CTACCAGGCA ATCTCAGCGG TATCAATCCC AACCGCTATG TTTTATCA GGATATTCTT	6300
TGTCCGATTC TTGATCAACA CATGACACCT GAACAGGACA AACCGCACCT CGCTCAGGCT	6360
GCTGAGACGC TTGCTAACAT TAAAGAAAAA GCTGGAACT ATGCCTATCT CTTTGAACT	6420
CAGGCCCCAGT TGAATGCTAT TTTAAGTAGC AAAGTAGATG TGGGACGACG CATTCTCAG	6480
GCCTACCAAG CGGATGATAA AGAAAGTTTA CAACAAATCG CCAGACAAGA ATTACCAGAA	6540
CTTAGAAGCC AAATTGAAGA CTTCCATGCC CTCTTTAGCC ACCAATGGCT GAAAGAAAAAC	6600
AAGGCTTTTG GTTTGGATAC AGTTGACATC CGTATGGGCG GACTCTTGCA ACGCATCAAA	6660
CGAGCAGAAA GCCGTATCGA GGTATATCTG GCTGGTCAGC TTGACCGCAT CGACGAGCTG	6720
GAAGTTGAAA TCCTACCATT TACTGACTTC TACGCAGACA AGGATTTGCG AGCAACTACA	6780
GGCAACCAGT GGCATACCAT TGGCAGAGCG TCGACGATT ATACGACTTA ATATTCTTCG	6840
AAAATCTCTT CAAACCACGT CAGCTTCCAT CTGCAACCTC AAAACAGTGT TTTGAGCAAC	6900
CTGCAGCTAG CTTCTAGTT TGCTCTTTGA TTTTCATTGA GTATAAAAAC AAGAACACCT	6960
TGCTTGCGCG AGGGTGTTC GCGTGAAACA GAAGAATTAT CTGGTTTCAA ATGCTACAGT	7020
TAGACAAACT TATGATAAAA TAGCAGAAAG TGAATGTTTC CTAAGAGCAA TTGGAGGTAT	7080
TATGCTACAC TTAAATTAG TAAACAAGA AATAGAAGCT GAAAAGCCAG CATCTGTAGA	7140
AGCTTGGATC ATTTCCGTCA AATTTAAAA AGGTTGCTAC CGACATATAT AGATTCCAAA	7200
AACAAAAACG TTAGCGGAAC TAGCAGATGT GATTTTATGG ACTTTTGATT TTGCAATGA	7260
TCATGCTCAC GCATTTTCA TGGATAATGT TGAGTGGAGT CATGCAGATT CTTACTTTCC	7320
TAGCTTTGTT AGTGACGATG TTGAAGAACG TTACACAGAA AATGCTATC TGGATAGCCT	7380

416

AAGTGTCAAA CAAAAATTTA AGTTTATTTT CGACTTCGGT GATGATGGC GTTTTGAATG	7440
CCAAGTGCTG AGAGAAATCG AGACAGAGGA CGAAGAAGCT TATCTCGTAC GTTCGGTTGG	7500
AACGTCGCCA GAACAATATC CAGATTATGA TGGTTTGGAC TATGAAGAAT GGTAAAATG	7560
AAATCAGTCT GTGTAGGCTT AGTATTTCAA TAGACTTCCT GCAAACTAG AATCCTAGTT	7620
CATGATTGAT AATACCAGCA ATCAAATTCA TTCGTAATCC GAAGCGTTTA CGATGATTTT	7680
GATAGGTTGT TGAAAACATT TTAAACGTTT TTTACTTTGGC AAAGATGTTT TCAACCTTGC	7740
TTCTCTCCTT AGATAGCGCA TGGTTATAGG CTTTATCTTC AGCTGTTAGT GGCTTGAGTT	7800
TGCTGGATTT ACGTGAAGTT TGTGCTTGAG GACATATCTT CATGAGCCCT TGATAACCAC	7860
TGTCAGCCAA GATTTTACCA GCTTGTCCGA TATTTCTGCA ACTCATTTTG AACAACTTCA	7920
TATCATGACA ATAGTTCACA GTGATATCCA AAGAAACAAT TCTCCCTTGA CTTGTGACAA	7980
TCGCTTGAGC CTTCATAGCG TGAAATTTCT TTTTACCAGA ATCATTGCTT AATTCTTTTT	8040
TTAGGGCGAT TGATTTTAC TTCCGTCGCA TCAATCATTA CCGTGTCTTC AGAACTAAGA	8100
GGAGTTCTTG AAATCGTAAC ACCACTTTGA ACAAGAGTTA CTTCAACCCA TTGGCTCCGA	8160
CGGATTAAGT TGCTTTCGTG AATACCAAAA TCAGCCGCAA TTTCTTCATA AGTGCGGTAT	8220
TCTAGGCTTA ATTTAGGTTT TCGTCCACCT TTTGCGTGTT TAAGTTGATA AGCTGTTTTT	8280
AATACAGCTA ACATCTCTTT AAAAGTCGTG CGCTGAACAC CAACAAGACG CTTAAATCGT	8340
GTATCAGTTA ATTGTTTACT TGCTTCATAA TTTGCGCAGG AGTCTATTGA CTCTTTGGTA	8400
GGTGTCAATG TTTTCTTCAT CTATCCCGAG AATTATTTTC CCGCCATTTG TATTTGCAAA	8460
TGCTGAGTAG GTTTCCCGAGA AAGACTCTGG AAGATTGTTT TTAGCTTTTT TGTATTCTAA	8520
ATCAACCCCT TCAAATTTTA AGTCCATATT TTTCTTTTAC ATCTGTTTTT TGTGGTTCTG	8580
GTATTTGTTT AAGTTGAGTG ATAATATAGC GAATTGAATT TCGAGAGTTT TTAAGTCACTT	8640
AATTTCTTTT TTAACCC	8657

## (2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11384 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

TCTATTTTGG GTATAGACTT ACCTATAAAG AAAAATATCT ATACACTGCC TTACTAGCTA	60
TACTGAACGA GTCAACAAAA ACCGATATATA TTGATGATAT AAATACAGCA AGATTTTTTA	120

417

ACTTCTTTGG CAATGATATT CCTAATTCGT CTTTAAAAAA AATTGACTAT ATCGCACCTT	180
CAGAAATTGT TTCATTAGT ACGTACGTTT GACAACGTTT TAAAGTAATT CCTAAAATTT	240
TGGAACATAT ATTAAAATCA AGTTTTTTAT TAGAGAATAT AGATGTTTCT GTTACACTG	300
TAAATATTTT AGAAGATCAA TTAACAAAAC ATAGAACAAT CAAAATTAGT AAAAATAAC	360
TGGTTGATCT CATGTATAAA TACCTAACAA AACCACGCGC CTTGCCTGCT GATGGAAAAG	420
AAGGTACAAA TACATGAATA TCAAAGAAAA AATCAAAAAG AATGGCCAAA GAGTTTATTA	480
TGCTAGTGTT TATCTAGGCG TTGACCAACT AACGGGCAAA AAAGCCCGTA CAACTGTTAC	540
AGCAACCACT AAAAAGGGCG TTAAAGTAAA AGCGCGTGAT GCGATCAATA CTTTGGCTGC	600
TAATGGCTAT ACAGTTAAAG ACAAGCCGAC AATTACAACA TATAATGAGC TTGTAAAAGT	660
TTGGTGGGAT AGTTACAAGA ATACAGTTAA GCCAAATACT CGCCAATCCA TGGAGGGATT	720
GGTTAGAGTG CATTTATTGC CTGTATTTGG CGATTACAAG CTATCTAAAC TTAATACGCC	780
TATTCTTCAA CAGCAAGTAA ACAAATGGGC TGACAAGGCA AATAAAGGCG AAAAAGGGGC	840
ATTTGCTAAC TACTCTTTGC TCCATAACAT GAATAAGCGT ATTTTGAAAT ATGGCGTAGC	900
TATCCAGGTA ATACAATACA ACCCAGCTAA TGATGTCATC GTTCCACGCA AACAGCAAAA	960
AGAAAAGGCT GCTGTCAAAT ACTTAGACAA CAAAGAATTA AAACAGTTTC TTGATTATTT	1020
AGATGCTCTG GATCAATCAA ATTATGAGAA CTTATTTGAT GTTGTCTGT ATAAGACTTT	1080
ATTGGCCACT GOTTGCCGTA TTAGTGAGGC TCTGGCTCTT GAATGGTCTG ATATTGACCT	1140
AGAAAGCGGT GTTATCAGCA TCAATAAGAC ACTAAACCGC TATCAGGAAA TAAACTCACC	1200
TAAATCAAGC GCTGTTATC GTGATATACC AATAGACAAA GCCACATTAC TTTTACTGAA	1260
ACAATACAAA AACCGTCAAC AAATTCAGTC TTGGAAATTA GCGCGATCTG AAACAGTTGT	1320
ATTCTCTGTA TTTACGGAGA AATATGCTTA TGCTTGTAAC TTACGCAAAAC GCCTAAATAA	1380
GCATTTTGAT GCTGCTGGAG TAACTAACGT ATCATTTTCT GGTTCCTGCC ATACACATAC	1440
TACTATGATG CTCTATGCTC AGGTTAGCCC GAAAGATGTT CAGTATAGAT TAGGCCACTC	1500
TAATTTAATG ATCACTGAAA ATACTTACTG GCATACTAAC CAAGAGAATG CAAAAAAGC	1560
CGTCTCAAAT TATGAAACAG CTATCAACAA TTTATAAAAA ATAAGGGTGA CCCATTTCCG	1620
GGCTACCCTC TTACTATACC AAAAATTAGT AGGGGTAGTA AAAAGGGTAT TAAATTATAA	1680
AAAGCACTAA GGGAAAGCGC CCCAAAGTGC TTATTTCAAA GGCTTTATAG CCTATAATCA	1740
CATAAGAGA TTATTTTTTA AGGTTGTAGA ATGATTTCAA TCCACGATAT TCAGCTACTT	1800
CACCAAGTTG GTCTTCGATA CGAAGCAATT GGTGTATTT AGCGATGCGG TCTGTACGTG	1860



418

AAAGTGAACC	AGTCTTGATT	TGTCCTGCGT	TAGTTGCAAC	TGCAATATCA	GGGATTGTTG	1920
AATCTTCAGT	TTCACCTGAA	CGGTGTGATA	CAACAGCAGT	GTAACCAGCT	TCTTTAGCCA	1980
TTTCGATAGC	TTCAAAAGTT	TCAGTAAGAG	TACCGATTTG	GTAACTTTG	ATAAGGATTG	2040
AGTTAGCAGC	ACCTTCTTGG	ATACCACGTG	CAAGGTAGTC	AGTGTGTTGTT	ACGAAGAAGT	2100
CGTCACCAAC	AAGTTGTACT	TTCTTACCAA	GACGTTTCACT	AAGAGCTTTC	CAACCATCCC	2160
AGTCGTTTTC	ATCCATACCA	TCTTCAATAG	TGATGATTGG	GTATTTGTTA	ACCAATTCTT	2220
CAAGGTAGTC	GATTTGTTCT	GCAGATGTAC	GAACAGCAGC	ACCTTCACCT	TCAAATTTAG	2280
TGTAGTCGTA	AACTTTACGT	TCTTTATCGT	AGAATTCTGA	TGAAGCACAG	TCAAATCCGA	2340
TAAATACGTC	TTTACCTGGT	ACATATCCAG	CAGCTTCAAT	CGCAGCAAGG	ATAGTTTCAA	2400
CACCATCTTC	AGTTCCTTCG	AAACGAGGAG	CGAATCCACC	TTCGTCACCT	ACGGCAGTTT	2460
CCAAACCACG	TGATTTAAGG	ATTTTCTTAA	GAGCGTGGAA	GATTTTCACT	CCGTAACGAA	2520
GGGCTTCTTT	AAATGTTGGC	GCACCAACTG	GCAAGATCAT	GAAGCTTGG	AAAGCGATTG	2580
GAGCGTCAGA	GTGAGAACCA	CCGTTGATGA	TGTTTATCAT	TGGAGTTGGA	AGAACTTTAG	2640
TGTTGAATCC	ACCAAGATAG	CTGTAAAGTG	GGATTTCAAG	GATGTCAGCA	GCAGCACGAG	2700
CTACAGCGAT	AGACACACCG	AGGATTGCAT	TCCGACCCAA	TTTACCTTTG	TTAGGAGTAC	2760
CGTCAAGTGC	GATCATAGCA	CGGTCAATAG	CTTGTTGATC	ACGTACATCG	TAGCCAATGA	2820
TAGCTTCAGC	AATGATGTTG	TTTACGTTGT	CAACAGCTTT	TTGTGTACCA	AGACCACCGT	2880
AACGAGATTT	GTCACCGTCG	CGAAGTTCAA	CTGCTTCGTG	TTCACCAGTA	GAAGCTCCTG	2940
ATGGAACCAT	ACCACGTCCG	AAAGCACCTG	ATTCAGTGTA	AACTTCTACT	TCAAGTGTTC	3000
GGTTACCGCG	TGAGTCTAGG	ACTTCGCGAG	CGTAAACATC	AGTAATAATT	GACATTTTTT	3060
ACTCTCCTTA	TGAGTTAAAT	TTTTTACACC	TCTATAATAC	CTTAAAACCC	CTCCTTTTTT	3120
AAGAAAAAAC	GTTATCTTTG	TGCAACTTTT	CCTTAACTTT	ATAAAGTAAT	CGCTTCTTTT	3180
TGTCTGTTTT	ATTCTAACTT	TTATGATATA	CTGTTTTCA	GACAGATTTA	TCAAAACAAT	3240
TACTTGAAAA	AGCTCATGGT	GGGTTAAAAA	TAAATCCGGA	TGAGCAAAGA	CGCTATCTTG	3300
GTACTTTTGA	GGAAAGAGTT	CTTGGATATG	TAGATATTGA	CACAGCAAAT	AGCCCTCAGT	3360
TAGAAAAAGG	CTTTTTATTT	ATTTTAGAAA	ACCTTCAGGA	AAAAGCAGAG	CCACTATTTG	3420
TGAAGATTTT	ACCAACTATC	GAATTTGATA	AGCAAGTTTT	CTACTTAAAA	GAAGCAAAAAG	3480
AAACTGATAG	TCAAGCCACC	ATAGTATCTG	AAGAGCATA	TACTTCTCCT	TTTGGCCTGG	3540
TTATTCATAG	CAATGCACCA	GTTCAAGTAG	AAGAAAAAGA	CCTTCGACTT	GCTTTTCCAA	3600
AACTTTGGGA	AGTTAAAAAG	GAAGAACCAG	CCAAAACATC	CTTATGGAAG	AAATGGTTTA	3660

GCTAAATCTT GCACATATTT AATAAGTGCC CAATATTGGC AGCCGTGCGC TCCAGATAGA	3720
AACTGGCATT TTCAAACATA TCTTCTAAAG GTTCACTTTT CTCCAAAATA GAAAAGACAG	3780
CTTGATATTT TTCAAATGGT AGGGGAGGTA AATCTTCAGC AAGACTACCG CAAATAGCAA	3840
TAACAGGAAC TCCAACAGGG GTTCTTTTTG CAACACCTAT AGGCGCTTTC CCAGCAAAGC	3900
TTTGACTATC AAGTCTTCCT TCTCCAACAA CAACCAAGTC AGCATCTGAA ACTTTCTTAT	3960
CAAAGTTGAT TAAGTCCAAG CAGGTATCAA TTCCAGACAC GATACTTGCC TGAGCAAAGG	4020
CACACAAACC ACCAGCAAGG CCTCCACCTG CTCCTGCTCC TTAAATTTCT AATGTTGCAG	4080
GTGAGAATTT TTCATAAAAA TCTTGGATCG CCTGATCTAC GACTGCAAAC ATAGTCGGAT	4140
GTAGACCTTT TTGATTGCCA AAAGTGTAAG TCGCACCTTG ATGACCACAT AAGGGACTCA	4200
CGACATCTGC TAAAATATGA ATTTGAACAC CTTCAGGAAT TTTATAGCAA TTTTCTGTTG	4260
AAACAGAAGC TAAGTTTAAT AAGGATTGAC CGGAAGCAGG CAAGACATTT CCATCCCTAT	4320
CATAAAATTG ATAACCTAAA CCAGCAGCAA TCCCCAGTCC TCCATCATT CTGGCCGTGC	4380
CACCAACACC GATATAAATA TCTTTAATCC CTTTAGAGAT GAGATGAAGA ATCAACTCTC	4440
CAATACCACA AGTTTGGATT TGAAGTGGAT TTCGTTTCTC TAGCGGAATT TTTCCAAGAC	4500
CAACCAAGTC AGCTACTTCA AATAGTGCCA GTTCCCCCTT TTGAAAATAG CGCATGGCTT	4560
CTTTTGTGCC AAAAGGGTCT GTCACTTGGA TCCATTTTTC TTTTAGGTCA AGAGAATGTC	4620
GGATAGCATC TACAGTACCT TCTCCCCCAT CACCAACAGG GCAGAGGAGA CATTCTACAT	4680
CTGCTATCGA TTGTTGGAAG CCTCTTTTTA TTGCTTCAGC TACCTGTTGA GCTGTCAAGC	4740
TTCTCTTAAA CGAATCCGGT GCAATTACAA TCTTCATATT TTCCCTCATT CTAAACAGTC	4800
AATCAAAGGG AGAACTTCTA AAAAATCCCT CTTGTCAACA TGATGTGGTA TTTCTTTTTT	4860
GAGCACTTCT TTGGCAGAAA AGGCGATTCC TAACTTCGCC GACTTCAACA TTAATAGATT	4920
ATTAACCCCA TCACCGATTG CCACCGTTCT TTCTTTAGAA AGTTTLAGTT TCTTCTCCA	4980
TTTTTCCAGA GTCTCTTTTT TGACCTGGGG ACTTATAATT TGTCCAACTA ATTTCTCTGT	5040
TAAAAGACCT TCTTTGACTT CAAGCTAGTT GGCAGTGAAA TAGGCAATAC CAAGGGATT	5100
TGCTAATCTC TCCAATATT GGTGTAAATC CACCAGACAC CAGACCAACT AGGATGCCAT	5160
TCTTTTGGAG AATAGAGATG AACTCTGGGA CATTAGCGA TAGATGAATT GAGTTGAAGA	5220
CGTTATCAAA GACCAAAATA GGAAGACCTT CCAACAAGGA CACTCTTTTT CTAAACTGC	5280
TTTCAAAGAC CAACTCTCCT CGCATGCTC GACTTGTAAT CTGCGAAATT TCCGCCTCAT	5340
GACCTGCCTC TCTCCCTAAA AGATCAATCA CTTCTTCTAG GATTAAGGTT CCATCTACAT	5400

420

CCAAAACACA	CAAGCCTTTT	ACTTGAGACA	TCAGTTCTCC	TCTCTAAACA	GCCTAAAAAT	5460
CGTATGAAGT	CATCATACGA	TTTTATCTAT	TAATTAAC TA	AACTATGGTA	CAAGTCAAGG	5520
TATGACTTGC	AGGCTGTATC	CCATGAGAAG	TCAGTCTCCA	TAGCTTGTTT	TTGTAGGTTT	5580
CTCCAAATGT	CTGGATGGTT	TCTATACAAG	TCCAATGCTG	TTTGAAAGT	CCAATTTAAC	5640
CAATAAGGAG	ATAGATTGTC	AAAGCTAAAG	CCAGTACCGC	TTCTTCGAT	TGGATTGAAA	5700
GCGCGAACTG	TATCTCGCAA	GCCTCCAAC T	TCATGGACCA	ATGGCAAGGT	TCCATAACGC	5760
ATAGCCATCA	TTTGAGACAA	GCCACACGGT	TCAAAACGAC	TTGGCATGAG	GAAGAGGTCA	5820
CAAGCAGCGT	AGATTTCCTG	AGCAAGTTTG	ACATCAAAAG	TGATATTTGT	TGATAGCTTG	5880
TCTGGGTAAA	TCTGAGCAAA	CCATGAGAAA	GCTCCTTCAA	AGGCTGGATC	GCCAGTTCCC	5940
AAAAGAACAA	TCTGAACATC	TTCTTGCAAG	ATATGGTGAA	GACTTTGAC	CACCACATCA	6000
AAACCTTTTT	GACGTGTCAA	ACGAGAAACA	ATTCACCA	GTGGAACGTC	TGCTCTAACA	6060
GGCAAGCCAA	CTCTTTCTTG	CAATTTTGCC	TTATTTTGG	CTTTCCCAGA	CAAATCTTCC	6120
TGATTGAAAT	GATAGTCTAA	AAGAGCATCC	GTCTGAGGAT	TATAAAGATC	AGCATCAATC	6180
CCATTCACGA	TACCAGATAC	TTTACCAGAC	TCCATTTTAA	GAATCTGATC	CAAATTACAT	6240
CCAAACTGAC	TAGTCATAAT	TTCATGAGCA	TAGCTAGGTG	AAACGGTTGA	AACACGGTTC	6300
GCATAGAGAA	TACCTGCCTT	CATCCAGTTC	AGACAGTTGT	TCCATCGAAG	GGTGCCATCA	6360
GCGTAACGTT	CAAAGCCAAC	TCCAAACAAA	TCACCCAACA	TTCTTCTGA	AAATTGTCTT	6420
TGGAATTCTA	AATTATGAAT	GGTTAAAACT	GTTTCAATGT	CCTCATAGGC	TTGAATCCAA	6480
CGGTATTTTT	CCCTCAACAA	GAAAGGAATC	ATAGCTGTAT	GGTACTCATG	AACATGGAGA	6540
AGATCAGGAA	TAAAGTCAAT	CCTTTCCATA	GCCTCAATGG	CAGCCAGTTG	GAAAAAGGCA	6600
AAGCGTTCTC	CGTCATCAAA	ATCACCGTAA	ACATGACCAC	GGAAGAAATA	ATATTGATTG	6660
TCAATAAAGT	AGAAGGTTAC	ACCATTTAAT	ACTGTTTTCT	TAATTCACA	ATACTGTCTG	6720
CGCCAACCAA	CGCTCACCTC	AAAATGAAGC	ACATCTTCAA	TCTGATTTCC	AAATTTAGCT	6780
TCTACCATAT	CATAGTAGGG	TAAAATCACT	GCAACTTCGT	GCCCAGCTTT	TACCAGTGAT	6840
TTTGGAAGAG	CGCCAATGAC	GTCTCCCAAA	CCACCTGTTT	TTGAAAAGGG	TGCACCCTCT	6900
GCTGCTACAA	ATAAAATTTT	CATGAATGAA	TATCTCTGT	TACTTTAGCA	CCTTTCTTAA	6960
CCACAAC TGG	ATGTTCTGCA	GTTCTCGAA	TCACAACACC	ATGCTCAACT	TCAACCCCTT	7020
TGTCCAAGAT	AGCATATTCG	ACCTGAGCCC	CTTCTCCAAT	AACAACACGA	GGGAAGAGCA	7080
GGCTATCTTT	AACCAAGCTA	TCCTTATGGA	CATGAATATT	ACGTGATAGA	ACAGAATTAG	7140
CTACTTGACC	TTCAATAATA	CTACCAGAGG	CAAAC T GAGA	AGTGCTTACC	TTAGATGTAT	7200

421

TAGCATAGTA AGTTGGCTCT TCGTTTTTGA CCTTTGTATA AATCTTTTGG TTTGGTGAGA	7260
AAAGAGAATA GAATTTTTGT GATTCAAGCA TATCGATATT CGCTTGATAA TAAGATTTAA	7320
CAGAGTGAAT ATTGGCTAGA TAGCCCGTGT ACTCGTAGGC GAAAGCTCCC TCTTTTACAG	7380
CCAAATCCCG TAAACATAG CGCAATTTCT CTGGATGTTT <u>TTTTTTAGCT</u> TCTTCTTCCA	7440
AGTGTTCAAT CAACCAAGGT GTATCAACGA CAAAGATATC TGTAGACATA TTGAACGTTT	7500
CAGCTGTGTA CTTGCTATCA AAGAGTTTAT GAGAAAGAAC ATGGTCTGTT TCATCTACAT	7560
CCAAGATTGC ATTTACTTCT GAAATATCTT TCTTAGCTAG TTTTTTATAA ACTACAGTGA	7620
TAGGCTCTTT TGTGTACTA TGTAGGTGGA AAAGTTGGTT CAAATCAATG TTAATAAGAA	7680
CATCGCAGTT GAGGGCAACC GTTTGGTTTG AGCCAGAACG TTTCAAATAA GTAAGAAGCT	7740
GTTGGTAGTA TTCTTTTCCA ACTGTACTAC TTTCTACACG GGTATTGTAA ATTCCTAGAT	7800
AGTAATGGCT AAGAAGGGTT GATAAGCCCC ACTCGCGTCC TGAACGAATA TGGTCAAATA	7860
CTGAGCTGAT ATTATCCTGC TGGAAAATAC CAAAGACACT ACGAACACCT GCATTAGCAA	7920
GGCTTGAAAG TGGGAAGTCA ATCAAACGAT ATTTCCCACC AAATGGCAAA CTTGCTACTG	7980
GACGGTGGTC CGTCAATGTC GACATATTGT GAAAACCAAC TGTATTTCTT AAAATGGCAG	8040
AATATTTATC AATCTTCATC TGTGCTACC CCCACTACTT CATTATATCC TACAACCTGT	8100
ACTTCATCTG TTCCATCAAT TTCGACACCG TCAGAAATAA TCGCACCTTC ACCAATAATG	8160
GCACGTTTAA TCTTAGCTCC TTGACCAATG ATAGCTCCAC TCATGATAAC TGAATCAAGG	8220
ACTCCGCTC CTTCGCGAAC TTGCGCGCTT GTTGAAAGGA TAGAATGTTT AACAGTTCCA	8280
TCAACGAAAC ATCCGTCTAC AACTAATGAG TCTTCCACAT GAGCATTTGC CCCGAGGAAG	8340
TTTGGTGGTG AAATCAAGTT TCTTGAGTAA ATCTTCCATT GACGGTTACG ACTATCCAAG	8400
GCATTTTCTG GAGAAATATA CTCCATGTTT GCTTCCCAA GTGACTCAAT AGTACCAACA	8460
TCTTTCCAAT AACCCTAAA TTCGTAAGCA TAAACACTTT CACCTGACTC AAGGTAATTT	8520
GGAATGACAT TTTTACCAA GTCTGACATG CCAACCTTGC TCTTTTCAGC AGCGACTAAC	8580
ATATTACGAA GCGGTTGCCA ATCAAAAATG TAGATTCCCA TAGAAGCTTT TGTAGATTTA	8640
GGTTGAGCTG GTTTTCTTC AAATTCAACA ATACGATTGT TAGCATCTGT GTTCATGATA	8700
CCAAAACGGC TTGCTTCTTT AAGAGGGACG TCTAAAACCTG CTAAGTCAA GCTGGCATTG	8760
TTATCCTTAT GAGACTGGAG CATATCATCA TAGTCCATTT TGTAGATGTG ATCCCCAGAC	8820
AAAATCAAGA CATACTCAGG ATTGACACTG TCGATATAGT CGATATTTTG GTAAATAGCG	8880
TGACTAGTCC CCTCAAACCA ACGATTTCTT TCACTTGCAG AATAAGGTTG AAGAATAGAG	8940

422

ACACCTGAAT TAATACCGTC TAGTCCCCAG CTTGAACCAT TCCCAATATG GTTGTGAGA	9000
GCAAGTGGTT GATACTGTGT AACGACCCCA ACATTGTGAA TCCCTGAGTT GGCACAGTTT	9060
GATAGGGCAA AGTCAATGAT ACGGTAGCGC CCACCAAATT GCACAGCTGG TTTTGCATG	9120
CTTTGAGTGA GTTTACCGAG ACGAGTTCCT TGCCCACCAG CAAGAATCAA AGCTAACATT	9180
TCATTTTCA TTTTCTACTC CTTTTTGGTT TTTATTTGTG ACGGTTTTAG TAGATTTCAA	9240
GCGACGTTTG ATTTTCATA CACTTGCTCC CATAGCCGGT AGGGTAAAGG TTAAGGCTG	9300
CTCATAATCT TTCCATAGTC CTTCTTGCGT TTGAACAGTT TGATTATGTT CTTTCCAAAC	9360
GCCTCCCCAC TCTTCCAACT CAGTATTCCA TACTTCTTCG TAAATTCCTG CAACGGGTAG	9420
TCCGATTGTA AAATCTTTCC GCTCAACAGG TACCATATTA AAGATACAGA CTAACATTTT	9480
TCCCTTTTTA CCCTTACGAA TAAAGGAAAG AACACTCTGG TCTCGATTAT CCGCATCAAT	9540
GATTTCAATA CCATCATAGC TGGTATCAAT TTCCCACAGA CAGCGATGAT CTTTGTAATA	9600
CTGGTTTAGC TGAGAAGCGA AATACTTCAT CTTAGCATTC ATTGGGTCTT CTAGGTTAGA	9660
CCATTCCAAC TGTCTTCAG ATTTCCATTC TAGGAATTGA CCGTATTCGC TACCCATGAA	9720
GAGCAATTTT TTACCAGGGT GACAAATTTG GTACGTATAG AGATTGCGCA AGCCTGCGAA	9780
TTGATTGTAA CGATCTCCCC ACATCTTATG CATCATACTC TTCTTGCCAT GAACCACTTC	9840
ATCGTGCGAG AATGGCAAGA GATAATTCTC CTTGAAAACA TACATAAAGC TGAAAGTCAC	9900
CAGGTAAAG TCATATTTAC GATAGATCGG ATCTTCTTCG TAGAAACGGA GGATATCATT	9960
CATCCAGCCC ATGTTCCATT TGTAGTCAAA TCCTAGACCA CCAATCTCTT TCATTCCCGT	10020
AATCTTGATC GCAGACGAAC TTTCTTCTGC AATCATCATC ACATCTGGAT ATTCTAACTT	10080
AATAACCTCA TTCAAGCGCT GAAGGAAATA ATAACCTTCA TAGTTGAGAT TTCCGCCATC	10140
TTTATTAGGT GTCCATGGAG CATCATCATA GTCCAAATAG AGCATGTTGC TAACAGCATC	10200
CACACGAATA CCATCCAAAT GATAGACATC AATCCAATGC TTAATGCAAG AAATTAAGAA	10260
GGACTGGACT TCATTTTTTC CAAGGTCAAA ATTAAGGGCA CCCCACCAT GGTATGAGC	10320
CTTATTATGG TCTTGGTATT CAAAAGTCGG TGTCCTCATCA TAATAGGCTA AGGCATCATC	10380
GTTGATGGTA AAGTGACTGG TACCCAGTCC ACAATAACCC CAATATTATG GGTATGACAC	10440
TCCTCGACAA AATCTTGAAA CTCCTCTGGT CGGCCATAAG CATGCTCTAA AGCGAAGTAA	10500
CCCATAAGCT GATACCCCA ACTCAAGCCC AAAGGATGGG ACATCAAGGG CATAAACTCA	10560
ATATGAGTAT AGTTCATTTT AACGAGATAA GGAATGAGTT CATCCTTGAG CTGGGCAAAA	10620
CTATAAGGAC TGCCATCAGA ATTTCTTTTC CATGATCCAG CGTGAACCTC ATAAATATTG	10680
ACAGGACGCT CTTCAAAGCC CCAACGTTTT CTTCGTGCCA GCCAAAGTCC ATCCTTCCAT	10740

423

TTCTTCTCAG GAAGCTCTGT TACGATTGCC CCTGTTCCCTG GACGAGCCTC ATACCTGACA	10800
GCAAAAGGGT CAATCTTCAT CAGTTGATGA CCATTTTGAC GTGTGACATG ATATTTGTAA	10860
ATATGCCCTT CTTGAGCCAT ATTGGTAAAG ACTTCCCAGA CCCCAAAATC ATTTCTTACC	10920
ATTGGAATCT GATTTTCAAT CCAGTTGGTA AAATCACCAA CCAAGTGAAC AGCCTGAGCA	10980
TTAGGTGCCC AAACACGGAA GGTATAGCCA TGCTCTCCAT TTAGTTCTTC CCTATGTGCT	11040
CCTAGATAAT GTTGGAGATA AAAATTTTCA CCCGTCATAA AGGTTTTTAA TGCTTCTCTA	11100
TTATCCATAT ACTCCCCCTC TCCTGTAAGC GTTTTCTATG TTTTATTAT ACTACCTTTT	11160
TAGAGAAGAT TCAAGTAAAT TACTATACTT CTTTAATTAT TTTGAAAATC TACAACAAGT	11220
TCACCTACTC GTTCAATTGT AAATCAATAT TTTTCAAAA AATTGCGAAA ACGCCTTTCT	11280
TTTTCTACTA TAGTGAAATG AAATAAAACA TCGCAAATC GATTAAGGAA TTAATCTAA	11340
TTTCTAACAA TGTCTTAGAA ATCAAAGTGT ACTATTTTAA CTCC	11384

(2) INFORMATION FOR SEQ ID NO: 46:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7577 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

TGTTGATTG TTAGTAGACG TTGACCAACG TCCTTCGGCT GGAAAAGGAA TTCTCCTTAG	60
TTTCCAACAC GTTTTCGCCA TGTTGGTGC GACCATCTTG GTACCATTGA TTTTGGGAAT	120
GCCTGTATCT GTTGCCCTTT TTGCTTCAGG TGTTGGAACA CTCATCTACA TGATTGCTAC	180
TGGTTTTAAA GTTCCAGTTT ATCTAGGTTT TTCATTGGCC TTTATCACAG CTATGTCACT	240
GGCTATGAAA GAAATGGGGG GGGATGTATC TGCTGCCCAA ACAGGGGTTA TCTTGACTGG	300
TTTGGTCTAT GTCCTTGTTG CTACCAGCAT CCGATTGTA GGAACAAAAT GGATTGATAA	360
ACTCTTGCCA CCAATCATT TCGGTCCTAT GATCATCGTT ATCGGCTTG GACTTGCAGG	420
TTCAGCTGTT ACCAATGCAG GTCTTGTAGC AGACGGAAAT TGGAAAAATG CTCTGGTAGC	480
CGTTGTTACT TTCCTAATTG CTGCCTTTAT CAATACAAAA GGAAAAGGCT TCCTACGAAT	540
CATTCCATT CTCTTGCCA TTATCGGTGG TTACCTTTTC GCACTAACTC TTGGCTTGGT	600
TGACTTTACA CCAGTTCTTA AAGCCAACTG GTTCGAAATT CCTGGTTTCT ACTTGCCATT	660
TAGCACAGGT GGTGCCTTTA AAGAGTACAA TCTTTACTTT GGTCCAGAAG CCATCGCTAT	720

424

CTTGCCAATC GCTATCGTAA CAATTCTGA ACATATCGGA GACCATACTG TTTTGGGTCA	780
AATCTGTGGT CGTCAATCTT TAAAAGAACC AGGTCTTCAC CGTACTCTTC TTGGTGACGG	840
TATCGCAACT TCTGTTCTG CCTTCCTTGG TGGACCAGCC AATACAACTT ACGGAGAAAA	900
TACAGGGGTT ATCGGTATGA CTCGTATCGC TTCTGTCTCA GTTATCCGTA ACGCTGCCCT	960
CATCGCGATT GCCCTCAGCT TCCTTGGTAA ATTCACTGCC TTGATTTCAA CTATTCCAAA	1020
CGCTGTACTT GGTGGTATGT CAATCCTTCT CTATGGGGTT ATCGCCAGCA ATGGTTTGAA	1080
AGTCTTGATT AAAGAACGTG TTGATTTGCG TCAAATGCGA AACCTCATCA TCGCAAGTGC	1140
TATGTGGTT CTTGGACTTG GAGGAGCTAT CCTTAACTT GGTCCAGTTA CACTTTCAGG	1200
TACTGCCCTT TCAGCCATGA CAGGAATCAT CTTGAACCTG ATCTTGCCAT ACGAAAATAA	1260
AGACTAAGAG TCTAAATACA CCTAATCCAC TCAGACAGCT GAGTGGATTT TTCGTATACC	1320
ATAATAAAG TGTCTTAACA AAATTATTAA AATCAAAAAA CGTATAATAT CAGATATCT	1380
AAAACCTTGA TACTGTACGT TTTATCATAG AAATTTTAC TTTATTTCT CATCAAATGA	1440
GATTTGCATC AATCTCTTGT CTTACTTGCG TTTCTTCTTC GCTTTCTTCA TTTTGTTAGC	1500
CATACGTTTC ATGGACTGTT TCATGGCAAA TTCACCAATT TTACCTTTCA AACCGCCACC	1560
AAACATCTGG CTCATATCTG GCATTCTGC TCCTCCGAGA GCTGATAAGT CAGGCATACC	1620
GCCTTGTCCT ATCATTCCTT CAAGGGCAGA CATATCCATT CCTCCCATAT TTGGCATATT	1680
TTTAGGAAGG TTATTTGGAT TAATCCCAT TTGCTTCATC ATTTTATTCA TATCCCAGA	1740
CATAACACCC TGCATGAGCT GTTTAGCCTG GTTAAAGTCC TTGATGAATT TATTGACTTC	1800
GACGAATGTA TTTCCAGAAC CAGCAGCAAT ACGACGGCGA CGGCTTGGAT TTAACAAATC	1860
TGGGTTTTCA CGCTCTTCAG GTGTATCGA AGACACAATG GCACGTTTAC GAGCAATCTG	1920
GCGTTCATCC ACCTTCATGT TTTGAAGGGC TGGATTGTTG GCCATACCTG GAATCATCTT	1980
GAGCAAGTCT TCCATCGGCC CCATATTTTG CACCTGATCT AATTGATCGA TGAAATCATT	2040
AAAATCAAAG GTGTTTTCGC GCATCTTCTC AGCCATTTCA AGGGCTTTTT GTTCATCGTA	2100
TTCTTGAGAA GCTTTCTCAA TCAAAGTGAG CATATCCCC ATACCAAGGA TACGGCTAGA	2160
CATGCCGTCT GGGTGAAGG TTTCAATGTC CGTAATCTTT TCACCTGTAC CAGTGAACCT	2220
GATTGGTTTT CCAGTAATGT GACGAACAGA CAGAGCAGCA CCACCACGAG TATCGGCATC	2280
AATCTTGGTA AGGATGACCC CAGTCACTTC CAACTGAGCA TTAAACTCAC GCGCAACATT	2340
GGCTGCTTCC TGACCAATCA TAGCATCAAC GACAAGCAAG ATTTCAATTTG GTTGAGCCAA	2400
TGCTTTCACA TCACGAAGCT CATTCATGAG GAGCTCATCA ATCTGCAAAC GACCCGCAGT	2460
ATCAATCAAG ACATAGTCGT TATGATTAGT TTGGGCTTGC TCCAAACCTT GACGTACAAT	2520

425

CTCAACAGCT GGTACTTCTG TTCCAAGTGC AAAGACAGGC ACATCAATCT GTTGTCCTCAA	2580
GGTCTTAAGC TGGTCAATGG CAGCTGGACG ATAAATATCC GCCGCAATCA TCAAAGGACG	2640
AGCATTCTCT TCTTTCTTGA GTTTGTGGC CAATTACCA GCAAAGGTTG TTTTACCAGC	2700
CCCTTGTAAG CCAACCATCA TGATGATGGT TGGAATCTTA GGTGACTTGA TAAATTCTGC	2760
CGTATCAGAA CCTAAAACGG CTGTCAATTC CTCATCAACG ATTTTAATAA TCTGTTGCGC	2820
AGGATTAAAGT GTATCAATGA CCTCATGCCC GACTGCACGC TCACGAACCTT TCTTGATAAA	2880
GTCTTTTACA ACAGGCAAGG CAACGTCGGC CTCGAGCAAG GCCAAGCGAA TTTCTTTGGT	2940
TGCCTCTTGG ACATCAGATT CAGAGATTTT TCCTTTTTTA CGTAGATTTT TAAAGACGTT	3000
CTGCAAACGT TCTGTTAAAC TTTCAAATGC CATTTTCTT CCTCTTATTC TCTATTATCA	3060
ATGCTTGTTA AAATTTCTAT CTGCTCCTGC AGAAAGTCAT CCTTGGGATA GCGCTCCAAA	3120
ATCTGATCAA AAATCTGACT GCGGACAATA TAGTCCGAGT ACATGTGCAA TTTCATCTCA	3180
TAATCTTCCA GAATCTTTTC TGTTGCTTG ATATTGTCTAT AGACAGCCTG ACGACTGACA	3240
CCGAACCTCT CGGCAATTC AGCAAGGCTG TAATCATCAG CGTAGTAGAG CTCGATATAA	3300
TTCATTGCT TATCTGTCAA AAGCGCCGCA TAAATTCAA AGAGCGCATT CATACGATTG	3360
GTTTTTTCGA TTTCCATAAC TTTTATTATA CCAAAAATTA GCCTAATCTA CCACACTAGG	3420
AAGCCGATCC AAGAAGATAG ATAGCTAAAT TTGAAAAAGA CATGAGCCTA GCCCAAGTA	3480
ATTTCCAATT GATAGCTGGC AAAGGGATGT CCTCTTGAT TTTGTAGTTG ATAATCTAGT	3540
TCAATCTTTT GCCTATCAAC TTGATAATGG CTCGTTTGA TGATAAACTC CTGCATGCCC	3600
ATAGGTGTAG GAATATAGGC TAACTATCG CTATCCTTA GAAAGCGCAT AATGGTCTTG	3660
GGATTAGAAA ATCGGCTCAT CACAAGTTCT TGACCATGAA ATTTAATCAC TACTTTTCC	3720
TTTTCTCAT TATAGAAAAG CAGGTAGCTA TAATCTCCTT TTTTATGCAC TTCCACATCA	3780
TAAAGCTGGT CAATCACTTC CAACTGCTCA TCAAACTGAA TCGTATTTTC CATCCGAATC	3840
TTTACATCAG GCCCTCTTC TTGTCTCTG TCCTACTATT TTACCAAAAA GAGCAGGATT	3900
TTGCTATAAT GGTCAATGA ACGAAAAAGT ATTCCGTGAC CCTGTTTACA ACTACATCCA	3960
TGTCAATAAT CAAATCATCT ATGACTTGAT TAATACAAAA GAATTTCAGC GTTTGCGCCG	4020
GATCAAACAA CTGGGAACCT CCACTTATAC CTTCACGGT GGAGAACACA GTCCGTTCTC	4080
TCACGTCTA GGAGTCTATG AAATGACAG ACGCATCACA GAGATTTTCG AAGAAAAATA	4140
TCCTGAGGAA TGGAATCCTG CCGAGTCTCT CTGACCATG ACCGCTGCTC TCCTACACGA	4200
CCTTGGGCAT GGTGCCTACT CCCATACTTT TGAACATCTC TTTGATACAG ACCATGAAGC	4260



426

CATTACTCAG	GAGATTATTC	AAAACTCTGA	GACAGAGATT	CACCAAGTCC	TGCTACAAGT	4320
GGCACCTGAT	TTCCCAGAAA	AGGTGGCCAG	TGTCATTGAC	CATACCTATC	CTAATAAGCA	4380
GGTCGTGCAG	CTCATTTCTA	GTCAGATTGA	CGCAGATCGC	ATGGACTATC	TCTTGCGCGA	4440
CTCCTATTTT	ACAGGAGCAT	CCTATGGGGA	ATTTGACCTG	ACTCGAATCC	TCCGAGTCAT	4500
TCGTCTTATC	GAAAATGGTA	TCGCCTTTCA	GCGCAATGGC	ATGCACGCCA	TCGAAGACTA	4560
CGTCCTCAGT	CGCTACCAGA	TGTACATGCA	GGTTTATTTT	CACCCCGCAA	CACGCGCCAT	4620
GGAGATTCTC	CTACAGAATC	TTCTCAAACG	CGCCAAGGAA	CTCTATCCTG	AGGACAAGGA	4680
TTTCTTTGCC	CGAACTTCTC	CACACCTCCT	GCCTTTCTTC	GAAAAAATG	TGACCTTGAC	4740
TGACTATCTG	GCTCTGGATG	ATGGCGTGAT	GAATACCTAC	TTCCAGCTTT	GGATGACCAG	4800
TCCTGACAAG	ATTCTTGAG	ATTTATCGCA	TCGCTTTGTC	AACCGCAAGG	TCTTTAAATC	4860
CATTACCTTT	TCACAAGAGG	ACCAAGATCA	ACTTACTAGC	ATGAGAAAAT	TGGTTGAGGA	4920
TATCGGCTTT	GATCCCGACT	ACTACACTGC	CATTCATAAG	AACCTTGACC	TCCCTTATGA	4980
TATCTATCGT	CCCGAATCTG	AAAACCCACG	GACACAGATT	GAGATTTTAC	AAAAAATGG	5040
AGAACTGGCC	GAACTCTCTA	GCCTGTCTCC	TATCGTCCAA	TCCCTTGCTG	GCAGTCGCCA	5100
CGGAGATAAT	CGCTTTTATT	TTCCAAAAGA	AATGTTGGAC	CAAAACAGCA	TCTTTGCAAG	5160
CATTACCCAG	CAATTTTATC	ACTTGATTGA	GAACGATCAT	TTTACCCCAA	ATAAAAACTA	5220
GAAGAGGAAA	TTTATGAGTA	TTAAACTAAT	TGCCGTGAT	ATCGACGGAA	CCCTTGTCAA	5280
CAGCCAAAAG	GAAATCACTC	CTGAAGTTTT	TTCTGCCATC	CAAGATGCCA	AAGAAGCTGG	5340
TGTCAAAGTC	GTGATTGCAA	CTGGCCGCCC	TATCGCAGGC	GTTGCCAAAC	TTCTAGACGA	5400
CTTGCAATTG	AGAGACGAGG	GGGACTATGT	GGTAACCTTC	AACGGTGCCC	TTGTCCAAGA	5460
AACTGCTACA	GGACATGAGA	TTATCAGCGA	ATCCTTGACT	TATGAGGATT	ATCTAGATAT	5520
GGAAATCCTC	AGTCGCAAGC	TCGGTGTCCA	CATGCATGCC	ATTACCAAGG	ACGGTATCTA	5580
TACTGCAAAT	CGCAATATCG	GAAAATACAC	TGTACACGAA	TCAACCCCTC	TCAGCATGCC	5640
TATCTTCTAC	CGTACCCCTG	AAGAAATGGC	TGGCAAAGAA	ATTGTTAAAT	GTATGTTTAT	5700
CGATGAACCA	GAAATCTCTG	ATGCTGCGAT	TGAAAAAAT	CCAGCAGAAT	TTTACGAGCG	5760
CTACTCCATC	AACAAATCTG	CTCCTTTCTA	CCTCGAACTC	CTTAAAAAGA	ATGTAGACAA	5820
GGGTTTACGC	ATTACTCACT	TGGCTGAAAA	ACTCGGATTG	ACCAAAGATG	AAACCATGGC	5880
AATCGGTGAT	GAAGAAAATG	ACCGTGCCAT	GCTGGAAGTC	GTTGGAAACC	CCGTTGTCTAT	5940
GGAAAATGGA	AATCCAGAAA	TCAAAAAAAT	CGCCAAATAC	ATCACCAAAA	CAAATGACGA	6000
ATCCGGCCTT	GCCCATGCCA	TCCGAACATG	GGTACTGTAA	AAGTATCATT	TTTCAATAAG	6060

427

AATTGATTAG CAATAAAATC CAATGAATTT TTTTAGCAAA CTATTTAATT TAAAACAAAA	6120
TAATCATAAT AGAGACACAA ATTCTGATTG TAACAATTTT TACCTAAACG AATTAGAATG	6190
TGGCCTTACT CCTGGGCAAC TCATACTCAT AGATTGGACT CAAAAAACAG GGAGAAATTA	6240
TAATTTCCCA AGATATTTTA AATACTCTCT TCAAATTGAC CCTGAATCTA CACACAATCA	6300
ATTATACAAA TTAGGATACT TCACTAAAAA TAAGACTTTA TCATATCTTA CAGTAGTAGA	6360
ATTAAAAACT ATATTATCTA AACATAATTT AGCTACTTCT GGAAAAAAG CAGAATTAAT	6420
TACAAGAATA ATTAATAATG TTAACATTGA CAATTTAGAT ATTCCGTTTCG AATTTAAACT	6480
AACAAAAGAA GCACAAAATC TTATTATCGA ACATAGTGAC TATATCAAAG CATACTATGA	6540
TAAAGACATA ACTATGGAAG ATTATTGTAA AGAAAAAAC AATATCTCTT TTAAAGCAAC	6600
TTTTGGTGAT ATAAATGGA GTCTCTTAAA TAAACAAGCT CATAGGAATA CTGTATCAGG	6660
AGATTTTGGA TGCTTATCTA ACACACGAAA GGCTCAGGGA AGACATTTGG AACAGAAGG	6720
TAATATTTAA CATGCTTTAA TATATTACAT AGAATCTTTG ATAATTACTA TTTCAGGATT	6780
AGAAAACAAT TTTTCAGCCA CTGATTATCC AGTATATTAT CCCGATTCTA TACCTGACTA	6840
CTCACTAAAA CATATTCAAA CATTAAATGA ATCATTATCT GATGACGATT ATGATTTTGC	6900
TTTGTATGAA GCATTATTTT GCTTCTCAAT TTTGAATGCA AATCATTTTT TATCTAAGGA	6960
AGATATTGAC TATTTAAGAG TTAATTTACC TCGTTCCACT GCTGAAGAAA TAAACAATTA	7020
CTTAAAGAAA TATGAATGTT ATAGTCCTTT AAATAATTTA GAACTTGACG ATTTTGAATA	7080
AATCGACTAT ACAAACATTT ATATACTCGA TATAGTCTCA ATTTTATCTG ATGATTGCCC	7140
AAATTTTTC AATAAAAC GCATAATATT ATGGAGACAA TCCCCTATAT TATGCGTTCT	7200
TTTAATATCA AAGACTTTTT GACAACTTC TTTGATATCT AATTACATGC CCCCTGCAGG	7260
AATCGAACCT GCAACTACTC CTTAGGAGGG AGTTGTTATA TCCATTGAAC TAAGGAGCT	7320
AGATAAAAC TCTGCTAAAT GAGCAGAGTT TTTTAGTCGA ATTAACGACC GATTTCTTTG	7380
ATACGAGCTG CTTTACCTTG AAGAGCAGC AAGTAGTACA ATTTCCGACG ACGTACTTTA	7440
CCGTAACGAA CAACTTCGAT TTTTCAACA CGTGGAGTGT GGATTGGGAA GATACGCTCA	7500
ACACCTACAC CGTTAGAGAT TTTACGAACT GTGTAGTTTT CTGAGATTCC AGCACCTTTA	7560
CGTGCGATAA CAACACG	7577

(2) INFORMATION FOR SEQ ID NO: 47:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4945 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

428

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

CCTCGCTGAT GATTGGTGCT GTTTTATTTG CTGGTCCAGC CTTGGCTGAA GAAACTGCAG	60
TTCTTGAAAA TAGCGGAnCT AATACAGAGC TTGTTTCAGG AGAGAGTGAG CATTGCGACCA	120
ATGAAGCTGA TAAGCAGAAT GAAGGGGAAC ATGCTAGAGA AAACAAGCTA GAAAAGGCAG	180
AAGGAGTAGC GATAGCATCT GAAACTGCTT CGCCAGCAAG CAATGAAGCT GCAACTACTG	240
AAACTGCAGA AGCAGCTAGC GCAGCTAAAC CAGAGGAAAA AGCAAGTGAG GTGGTTGCAG	300
AAACACCATC TGCAGAAGCA AAACCTAAGT CTGACAAGGA AACAGAAGCA AAGCCCGAAG	360
CAACTAACCA AGGGGATGAG TCTAAACCAG CAGCAGAAGC TAATAAGACT GAAAAAGAAG	420
TCCAGCCAGA TGTCCTTAAA AATACAGAAA AAACATTAAA ACCAAAGGAA ATCAAATTTA	480
ATTCTTGCGA AGAATTGTTA AAATGGGAAC CAGGTGCTCG TGAAGATGAT GCTATTAACC	540
GCGGATCTGT TGTCCTCGCT TCACGTGCGA CAGGTCATTT AGTCAATGAA AAAGCTAGCA	600
AGGAAGCAAA AGTTCAAGCC TTATCAAACA CCAATTCTAA AGCAAAAGAC CATGCTTCTG	660
TTGGTGGAGA AGAGTTCAAG GCCTATGCTT TTGACTATTG GCAATATCTA GATTCAATGG	720
TCTTCTGGGA AGGTCTCGTA CCAACTCCTG ACGTTATTGA TGCAGGTCAC CGTAACGGGG	780
TTCTGTATA CGGTACACTC TTCTTCAACT GGTCTAATAG TATTGCAGAT CAAGAAAGAT	840
TTGCTGAAGC TTTGAAGCAA GACGCAGATG GTAGCTTCCC AATTGCCCCG AAATTGGTAG	900
ACATGGCCAA GTATTATGGC TATGATGGCT ATTTCATCAA CCAAGAAACA ACTGGAGATT	960
TGGTTAAACC TCTTGGAGAA AAGATGCGCC AGTTTATGCT CTATAGCAAG GAATATGCTG	1020
CTAAGGTAAA CCATCCAATC AAGTATTCTT GGTACGATGC CATGACCTAT AACTATGGAC	1080
GTTATCATCA AGATGGTTTG GGAGAATACA ACTACCAATT CATGCAACCA GAAGGAGATA	1140
AGGTTCGGC AGATAACTTC TTTGCTAACT TTAAGTGGGA TAAGGCTAAA AATGATTACA	1200
CTATTGCAAC TGCCAACCTG ATTGGTCGTA ATCCTTATGA TGTATTTGCA GGTTTGGAAT	1260
TGCAACAGGG TGGTTCCTAC AAGACAAAGG TTAAGTGGAA TGACATTTTA GACGAAAATC	1320
GGAAATTGCG CCTTCTCTT GGTTTATTTG CCCAGATAC CATTACAAGT TTAGGAAAAA	1380
CTGGTGAAGA TTATCATAAA AATGAAGATA TCTTCTTTAC AGGTATATCAA GGAGACCCTA	1440
CTGGCCAAAA ACCAGGTGAC AAAGATTGGT ATGGTATTGC TAACCTAGTT GCGGACCGTA	1500
CGCCAGCGGT AGGTAATACT TTTACTACTT CTTTAAATAC AGGTCATGGT AAAAAATGGT	1560
TCGTAGATGG TAAGGTTTCT AAGGATTCTG AGTGAATTA TCGTTCAGTA TCAGGTGTTT	1620

429

TTCCAACATG GCGCTGGTGG CAGACTTCAA CAGGGGAAAA ACTTCGTGCA GAATATGATT	1680
TTACAGATGC CTATAATGGC GGAAATTCCC TTAAATTCTC TGGTGATGTA GCCGGTAAGA	1740
CAGATCAGGA TGTGAGACTT TATTCTACTA AGTTAGAAGT AACTGAGAAG ACCAAACTTC	1800
GTGTTGCCCA CAAGGGAGGA AAAGGTCTTA AAGTTTATAT GGCATTCTCT ACAACTCCAG	1860
ACTACAAATT CGATGATGCA GATGCATGGA AAGAGCTAAC CCTTTCTGAC AACTGGACAA	1920
ATGAAGAATT TGATCTTAGC TCACTAGCGG GTAAAACCAT CTATGCAGTC AAACATATTT	1980
TCGAGCATGA AGGTGCTGTA AAAGATTATC AGTTTAACCT AGGACAATTA ACTATCTCGG	2040
ACAATCACCA AGAGCCACAA TCGCCGACAA GCTTTTCTGT AGTGAAACAA TCTCTTAAAA	2100
ATGCCCCAGA AGCGGAAGCA GTTGTGCAAT TTAAAGGCAA CAAGGATGCA GATTTCTATG	2160
AAGTTTATGA AAAAGATGGA GACAGCTGGA AATTACTAAC TGGCTCATCT TCTACAACTA	2220
TTTATCTACC AAAAGTTAGC CGCTCAGCAA GTGCTCAGGG TACAACTCAA GAAGTGAAGG	2280
TTGTAGCAGT CGGTAAAAAT GGAGTTCGTT CAGAAGCTGC AACCACAACC TTTGATTGGG	2340
GTATGACTGT AAAAGATACC AGCCTACCAA AACCCTAGC TGAAATATC GTTCCAGGTG	2400
CAACAGTTAT TGATAGTACT TTCCCTAAGA CTGAAGGTGG AGAAGGTATT GAAGGTATGT	2460
TGAACGGTAC CATTACTAGC TTGTCAGATA AATGGTCTTC AGCTCAGTTG AGTGGTAGTG	2520
TGGATATTCG TTTGACCAAG CCACGTACCG TTGTTAGATG GGTCATGGAT CATGCACGAG	2580
CTGGTGGTGA GTCTGTAAAC GATGGCTTGA TGAACACTAA AGACTTTGAC CTTTATTATA	2640
AAGATGCAGA TGGTGAGTGG AAGCTAGCTA AGGAAGTCCG TGGTAACAAA GCACACGTGA	2700
CAGATATCAC TCTTGATAAA CCAATCACTG CTCAAGACTG GCGCTTGAAT GTTGCTACTT	2760
CTGACAATGG AACTCCATGG AAGGCTATTC GTATCTATAA CTGGAATAATG TATGAAAAGC	2820
TTGATACTGA GAGTGTCAAT ATTCCGATGG CCAAGGCTGC AGCCCGTTCT CTAGGCAATA	2880
ACAAGGTACA AGTTGGCTTT GCAGATGTAC CGGCTGGAGC AACTATTACC GTTTATGATA	2940
ATCCAAATTC TCAAACCTCC CTGCAACCT TGAAGAGCGA AGTTGGAGGA GACCTAGCAA	3000
GTGCACCATT GGATTTGACA AATCAATCTG GTCTTCTTTA TTATCGTACC CAGTTGCCAG	3060
GCAAGGAAAT TAGTAATGTC CTAGCAGTTT CCGTTCCAAA AGATGACAGA AGAATCAAGT	3120
CAGTCAGCCT AGAAACAGGA CCTAAGAAAA CAAGCTACGC CGAAGGGGAG GATTTGGACC	3180
TTAGAGGTGG TGTCTTCGA GTTCAGTATG AAGGAGGAAC TGAGGACGAA CTCATTGCCC	3240
TAACTCACGC AGGTGTATCA GTATCAGGTT TTGATACGCA TCATAAGGGA GAACAGAATC	3300
TTACTCTCCA ATATTTGGGA CAACCGGTAA ATGCTAATTT GTCAGTGACT GTCAGTGGCC	3360

430

AAGACGAAGC AAGTCCGAAA ACTATTTTGG GAATTGAAGT AAGTCTGAA CCGAAAAAAG	3420
ATTACCTAGT TGGTGATAGC TTAGACTTGT CTGAAGGACG CTTTGCAGTG GCTTATAGCA	3480
ATGACACCAT GGAAGAACAT TCCTTTACTG ATGAGGGAGT TGAATTTTCT GGTTACGATG	3540
CTCAAAAGAC TGGTCGTCAA ACCTTGACGC TTCATTACCA AGGCCATGAA GTTAGCTTTG	3600
ATGTTTTTGGT ATCTCCAAAA GCAGCATTGA ACGATGAGTA CCTCAAACAA AAATTAGCAG	3660
AAGTTGAAGC TGCTAAGAAC AAGGTGGTCT ATAACTTTGC TTCATCAGAA GTAAAAGAAG	3720
CCTTCTTGAA AGCAATTGAA GCGGCCGAAC AAGTGTGAA AGACCATGAA ACTAGCACCC	3780
AAGATCAAGT CAATGACCGA CTTAATAAAT TGACAGAAGC TCATAAAGCT CTGAATGGTC	3840
AAGAGAAATT TACGGAAGAA AAGACAGAGC TTGATCGCTT AACAGGTGAG GTTCAAGAAC	3900
TCTTGGCTGC CAAACCAAAC CATCCTTCAG GTTCTGCCCT AGCTCCGCTT CTTGAGAAAA	3960
ACAAGGCCTT GGTTGAAAAA GTAGATTGA GTCCAGAAGA GCTTACAACA GCGAAACAGA	4020
GTCTAAAAGA TCTGGTTGCT TTATTGAAAG AAGACAAGCC AGCAGTCTTT TCTGATAGTA	4080
AAACAGGTGT TGAAGTACAC TTCTCAAATA AAGAGAAGAC TGTCATCAAG GGTTTGAAAG	4140
TAGAGCGTGT TCAAGCAAGT GCTGAAGAGA AGAAATACTT TGCTGGAGAA GATGCTCATG	4200
TCTTTGAAAT AGAAGGTTTG GATGAAAAAG GTCAAGATGT TGATCTCTCT TATGCTTCTA	4260
TTGTGAAAT CCCAATTGAA AAACATAAGA AAGTTAAGAA AGTATTTTTC TTACCTGAAG	4320
GCAAAGAGGC AGTAGAATTG GCTTTTGAAC AAACGGATAG TCATGTTATC TTTACAGCAC	4380
CTCACTTTAC TCATTATGCC TTTGTTTATG AATCTGCTGA AAAACCACAA CCTGCTAAAC	4440
CAGCACCACA AAACACAGTC CTTCCAAAAC CTACTTATCA ACCGACTTCT GATCAACAAA	4500
AGGCTCCTAA ATTGGAAGTT CAAGAGGAAA AGGTTGCCTT TCATCGTCAA GAGCATGAAA	4560
ATACTGAGAT GCTAGTTGGG GAACAACGAG TCATCATACA GGGACGAGAT GGACTGTTAA	4620
GACATGTCTT TGAAGTTGAT GAAAACGGTC AGCGTCGTCT TCGTTCAACA GAAGTCATCC	4680
AAGAAGCGAT TCCAGAAATT GTTGAAATTG GAACAAAAGT AAAACAGTA CCAGCAGTAG	4740
TAGCTACACA GGAAAAACCA GCTCAAATA CAGCAGTTAA ATCAGAAGAA GCAAGCAAAC	4800
AATTGCCAAA TACAGGAACA GCTGATGCTA ATGAAGCCCT AATAGCAGGC TTAGCCAGCC	4860
TTGGTCTTGC TAGTTTAGCC TTGACCTTGA GACGGAAAAG AGAAGATAAA GATTAAATAT	4920
CGAAAAATCT TGTGAAATCT TTCCG	4945

(2) INFORMATION FOR SEQ ID NO: 48:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25002 base pairs
  - (B) TYPE: nucleic acid

431

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

GACAACTCAA GTAGCTTTTT CTTATTTTGA AAAAGGAGAT CAGAGTTTAA CTATGTCAGA	60
AAAATCACAA TGGGGGTCGA AACTTGCTTT TATTCTAGCA TCTGCTGGCT GGCCATCGGG	120
CTTGGTTCCG TTTGGAAGTT TCCCTACATG ACTGCTGCTA ATGGCGGTGG AGGCTTTTTA	180
CTAATCTTTC TCATTTCAC TATTTTAATC GGTTCCTC TCCTGCTGGC TGAGTTTGCC	240
CTTGCCCGTA GTGCTGGCGT TTCCGCTATC AAAACCTTTG GAAAACTGGG CAAGAATAAC	300
AAGTACAACT TTATCGGTTG GATTGGCGCC TTGCCCCTCT TTATCCTCTT ATCTTTTTAC	360
AGTGTATCG GAGGATGGAT TCTAGTCTAT CTAGGTATTG AGTTTGGGAA ATTGTTCCAA	420
CTTGGTGGAA CGGGTGATTA TGCTCAGTTA TTTACTTCAA TCATTTCAAA TCCAGCCATT	480
GCCCTAGGAG CTCAAGCGGC CTTTATCCTA TTGAATATCT TCATTGTATC ACGTGGGGTT	540
CAAAAAGGGA TTGAAAGAGC TTCGAAAGTC ATGATGCCCC TGCTCTTTAT CGTCTTTGTT	600
TTTATCATCG GTCGCTCTCT CAGTTTGCCA AATGCCATGG AAGGGGTTCT TTAATTCCTC	660
AAACCAGACT TTTCAAAACT GACTAGCACT GGTCTCCTCT ATGCTCTGGG ACAATCTTTC	720
TTTGCCCTCT CACTAGGGGT TACAGTCATG TTGACCTATG CTTCTTACTT AGACAAGAAA	780
ACCAATCTAG TCCAGTCAGG AATCTCCATC GTAGCCATGA ATATCTCGAT ATCCATCATG	840
GCAGGTCTAG CCATTTTCCA AGCTCGATCC CCCTTCAATA TCCAGTCTGA AGGGGGACCC	900
AGCCTGCTCT TTATCGTCTT GCCTCAACTC TTTGACAAGA TGCCTTTTGG AACCATTTTC	960
TACGTCCTCT TCCTCTTGCT CTTCCTTTTT GCGACAGTCA CTTTTTCTGT CGTGATGCTG	1020
GAAATCAATG TAGACAATAT CACCAACCAG GATAACAGCA AACGTGCCAA ATGGAGTGTT	1080
ATTTTAGGAA TTTTGACCTT TGCTTTTGGC ATTCCTTCAG CCCTATCTTA CGGTGTCATG	1140
GCGGATGTTT ACATTTTGG TAAGACCTTC TTTGACGCTA TGGACTTCTT GGTTTCCAAT	1200
CTCCTCATGC CATTTGGAGC TCTCTACCTT TCACTTTTTA CAGGCTATAT CTTTAAAAAG	1260
GCTCTTGCAA TGGAGGAACCT CCATCTCGAT GAAAGAGCAT GGAAACAAGG ACTGTTCCAA	1320
GTCTGGCTCT TCCTTCTTCG TTTCTTCGTT TCGTCATTCC AATCATCATC ATTGTGGTCT	1380
TCATTGCCCA ATTTATGTAA TCAAAAAGGA CTTGAGTAGT GAACTCAGGC CTTTCTTTT	1440
TATGGATGGC TAACAATCAA TTCCAAACCT TGCCCTTCCA GAGTCCAAGC TTCAACATCA	1500
CTTGGTAGGA TAAAGTGGCT GCCTTTTTGA ATTGGATAAT TTTTCCCCTC AACAGTTAGC	1560

432

TGACCTTGAC CAGCCAAGAC ACTCAATAAG CTGTAGTCAG CTGTCTTTTC AAAGTCAACT	1620
TTTCCAGTAA TTTCCCACTT GTAAACTGCG AAGAAATCAT TAGATACAAG GAGAGTGGAA	1680
CGCAAATCAT CTGCTTTAAC AGTTACAGGA CGGCTATTTG CTGGCTCACC AATGTTCAAG	1740
ACATCGATGG ATTTTTC AAG ATGAAGTTCA CGCAAGTTGC CTTTGTCTATC CTTGCGGTCA	1800
AAGTCATAGA CGCGATAGGT GGTATCGCTA GACTGCTGGG TTTCAAGGAT TAAGATACCC	1860
GCCCCGATAG CGTGCATAGT CCCGCTTGGT ACATAGAAGA AATCTCCAGC CTTAACAGGG	1920
ACTTTGGTCA ACAAGTCATC CCAGTTCTTG TCCTCGATTT GCTGGCGGAG TTCTTCTTTT	1980
GACTTGGCAT TGTGACCGTA GATAATCTCT GAACCTTCAT CCGCTGGGAT AATGTACCAG	2040
CATTCTGTTT TTCCGAGTTC GCCTTCATGC TCGAGTCCAT AAGCATCGTC TGGGTGAACT	2100
TGGACACTGA GCCAGTCGTT GGCATCGAGG ATCTTGGTCA AAAGTGGAAA TACAGTTCTT	2160
GGACGATTGC CAAATAATTC ACGGTGTTCC GCATACAAAG TAGCAAGATC TGTTCCTCTG	2220
TAACGACCAT TGGCAACTTT AGAGACTCCA TTTGGATGGG CTGAGATGGC CCAATATTCT	2280
CCGATTTTTT CACTTGGGAT GTCGTAGCCA AACTCATCAC GTAGCTTGGC TCCACCCCAG	2340
ATTTTTTCTT GCATAACTGA TTGTAAAAAT AATGGTTCTG ACATGTCGAT CTCCTGTCTG	2400
ATTTTTCTCC CCTCATTATA GCAAAAAAAG AGTTCGAATT GAACTCTTTT TTACATCTTA	2460
TAAAGCAGGG AGAAGATTTT ATAAAAATAG TAAACAAATG TGCTCTACCC GATGCTTGCA	2520
CCATTGCTAT AAATGACATC CTGTGTACCA TAGAAGGACT TCTTCTTGCT ACGTTTGAGA	2580
GCTCCGTTTC CTACATTATC TCGATCTACA TAGATAAAGC CATAGCGCTT ATTCAATTCC	2640
CCTGTGCCAG CTGAAACCGG ATCGATACAG CCCCAGTCTG TATAACCAAG CAAGTCAACC	2700
CCGTCTTGGT AAATGGCATC TCGCATGGCC TTGATGTGGG CCTCTAAGTA AGTAATCCGA	2760
TAGTCATCTG CTACATAACC ATTCTCATCC GGTGTATCCA TAGCACCAGG TCCATTTTCT	2820
ACGATAATAC TAACTAAAA TCAAAAAGCA TTATATAATA GTGATATGAA ATCAACTAAA	2880
GAAGAAATCC AAACCATCAA AACACTTTTA AAAGACTCTC GTACAGCTAA ATATCATATA	2940
CGCCTTCAAA TCGTTCTATA GTAAAATGAA ATAAGAACAG TACAAATCGA TCAGGACAGT	3000
CAAAATCGATT TCTAACAATG TTTTAGAAGT AGGGGTGTAC TATTCTAGTT TCAATCTACT	3060
ATATTTCTGC TGATGGGCAA ATCTTATAAA GAGATTATAG AACTTTTATA GTAGTTTGAA	3120
ATAAGATGTG AACAACTCTA TCAGGAAAGT CAAATTAATT TATAGAAATA TTTTAGCAGC	3180
CAAGGTGTAC TGTATAGAT TCAATACACT ATAGACTGTA ATCAAACAAC GATTTGGCGA	3240
AATGTAAAAA AATATGAGGA GTTCGGACTC GACTCTCTCC TTCAAGAAAC ACGTGGTGGT	3300
CGTAACCATG CATATATGAC AGTTGAGGAA GAGAAAGCCT TTCTTGCCCG CCATTTGAAG	3360

433

GCTACAGAGG CAGGAGAATT TGTTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG	3420
TTAGGTCGTT CCTACACACG TGATGCCCTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA	3480
AATATTACGC CACGTCCAGA ACATCCTAAG AAAGCAGACG CTCAAACCAT TGTGCGTCT	3540
AAAAATAAAA TCTCAATCCA AGAAGGCAAG AAAGCGTTT AAATATAGTA GACGTTTTCG	3600
TAAGGTTTGC TTGATGTACC AAGCTGAAGC TGGTTTCGGT AGAATCAGTA AACTGGGATC	3660
TTGTTGGGCT CCAATAGGAG TAGGTCCACA TATCCATAGT CACTATATAC GAGAATTTTCG	3720
CTATTGTTAT GGAGCTGTTG ATGCCTATAC AGGCGAATCA TTTTCTTAA TAGCTGGTAG	3780
ATGTAATACT GAGTGGATGA ACGCCTTTT AGAAGAGCTT TCACAAGCTT ATCCTTTTAC	3840
TCGTTATGGA CAATGCTATA TGGCATAAAT CAAGTACCTT AAAGATTCCG ACTAATATTG	3900
GTTTTGCATT TATCCTCCA TACACACCAG AGATGAACCC CATTGAACAA GTGTGGAAG	3960
AGATTTCGTAA ACGTGGATTT AAGAATAAAG CCTTTCGAAT TTTGGAAGAT GTCATGAATC	4020
AACTCCAAGA TGTCATACAA GGATTGGAGA AGGAGGTGAT AAAGTCCATC GTTAATCGGA	4080
GATGGACTAG AATGCTTTTT GAAAGCAGAT GAGTATTATA TGCAATTTCT TTATATAAAA	4140
AGACCGGATT GCTCCGATCT TTCAATAGTT CATATTCTCA ATTTCTATTT TAAAAATAGC	4200
TAAGGTAAAC GTCAAATGAC TACGCGACCT ATTTCATACG ATAAAAATCA AGCACTAGAC	4260
CAGCAGGTCC TTGAACTAAT AAGGACTCTG TCCCCAATC GGTTACAGTT GGTCCGTGTA	4320
AAACCTTTAT ACCAAGCTCG TTCAACCGTT TGTAGTTCTG GTCTACATCC TCAACCTCGA	4380
TATGAATAAT GATTCTTGAC TGAAAGTTTT CCAAAGGAAC CAAATGATTT TGTGACAACA	4440
TAAGGCAGTG ACTACCAATC GTAAACTGAG CAAAACCATC ATTAGCATAA TCTGCCTTTT	4500
TATCCAAGAT ATGCTCCAAG TCAGCACAGA CTGGGGAAC ATTTGAAACG ATAATATCTA	4560
ATTGATTTAA ATTCATTTAC TCTCCTCCAT AAAAAGACCG GATTGCTCCG ATCTTTTAAA	4620
GTTCTGCTCT ATGAAAATCA AAGAATAAAG TCTACAAGTT TCATATTTGA TTTTCGGCGA	4680
GAGGAATTAT TTAATTGCGC GTGATTGCAA TCCTTCTTCT TCCAAGAAGA GACGGAATGG	4740
TACGAGTTCT TCTGCTTCGT ATTTTTCCTT GAAGGCTTTG ATAGCTTCTT CTGAGTGAAG	4800
TTTTGGATCC AATTCAAGTA CTTCTACTGG AAGTGGACGG TGTGAGTGA TGCGAGCATC	4860
GATGACAACA GTTTTACCTT CTTTGTTCAA TTTAACAGCT TCTGCAACAA CTGCATCGAT	4920
GTCTTCGATA CGGTCAACTG TGAATCCAAC AGCTCCTTGA GCTTCCGCAA TTTTAGCGTA	4980
GTCAGCGTTT GTGAAGTCTA CACCAAACAA GTGTTTGTG GTATCTTCGT ATTTGTTCTT	5040
GATGAAGCCG TACTCAGCAT TTGAGAAGAC AAGGTTGATA ACTGGAAGGT CGTATTGAAC	5100



434

GTTCGTGATA ACGTCTGGGT AGCACATGTT GAATGCTCCG TCACCCATGA TGTTCATAC	5160
TTGGCGATCT GGATTGTCTT TCTTAGCAGC GATACCACCA GGAAGGGCAA TACCCATTGT	5220
CGCAAAGAGT GGAGATGTAC GCCACATGTT CTTAGGTGTC ATGTGAAGGT GACGAGTAGA	5280
TGTTTGAGTA GTGTTACCTA CGTCGATTGA GTAGATAGCG TCTTGATCAG CATGTTTGTT	5340
GATTGCATTG TAAACTTGAT ACAATTGCAA TTCACCCTCA GTTTTACCTT CGAGTTTGTT	5400
CATGTAATCA CGCCAGTTTT GGTGTTCTT AACGTTTGCA CGCCACCATG GAGTTGATTC	5460
AACTGGGTTT ACTTTGTCAA GGATAGCTTT AGCTGCTTGA CCAGCATCAC CAAGGATTGA	5520
AGCGTCAAGG GCATGACGTT TACCAAGTTT GTAAGGGTCG ATATCGACTT GGATGAATTT	5580
TTCAGTGTTT TTGAATGCTT CGTAACTTC AGCAAATGGG AAGTTTGAACT CAAGGAAAAG	5640
AACTGTGTCT GCTTCAAAGA CCACTTCGTT GGCTGGTTTC CAACCAACAC GGTAAAGCAGA	5700
ACCTGTCAAA CCTTCATAGT TCCATTCGAA AGCTTCAAAG TTTTACCAG TTGTGATGAT	5760
TGGTGCTTTG ATTTTACGTG ACAATTGAGT AATCACTTCA CCAGCTTTAA CACCACCAA	5820
TCCAGCATAG ATAAGTGGG GTTCAGCATT GTTCAAGATT TCAACAGCTT TGTCGATTTT	5880
AACTTCGTTT AAAGCAGGAG CGATGAATGA GCGTTCGTAT GAACCTGAAC CGTAGTATGA	5940
GTTCCTCATG ATTTCTTGA AACCAGAGTT TACTGGAATT TCAACAACAG CTGGACCTTT	6000
TTTAGAAACT GCAGCACGGC AGGCTTCGTC AATTACTTTT GGCAATTGCT CAGCGTAAGC	6060
TACACGTTTG TTGTAACAG CGATACCGTT GTACATTGGG TTTTGGTTAA GCTCTTGAA	6120
AGCATCCATG TTCAATTCGT TAACTGGAGC TGATCCAAGG ATCGCTAGGA ATGGAGTGTT	6180
ATCCATAGCT GCATCGTAAA CACCGTTAAT CAAGTGAGTC GCACCTGGAC CACCTGAACC	6240
AACTGCAACC CCGATTGAGC CGCCGAATTT AGCTTGCTATA ACCGCTGCAA GAGCACCTGT	6300
CTCTTCGTGG CGAACTTGTA AGAAACGGAT ATCTTTGTCT TCAGCCAAAG CGTCCATCAA	6360
TGAGCTGAGT GTTCCTGATG GGATACCGTA GATTGTATCT ACGCCCCATG TTTTCAATAC	6420
GTTAAGCATT GCTGCAGATG CAGTAATTTT CCTTGAGTC ATAATGATAA CTCTCCTTCA	6480
ATTTTTTTAA ACTTGAGAA TACGATTACA TAGAATTGGA AACGTTCTCC AAATTTTAC	6540
TATTCCTAGT TATCATATTT ATGCTGACTT TTCTAAAAAT CTGCTCAAAA CTCTCTATTC	6600
TCTATTCTAA TACAGTTTTG AAAGTTCTGT CATTTCTGTT TTATAACAAA GAAATCTAGT	6660
CATTACTTTT AGTCTATTTT ACTAAAAAT AACAGAAGGG AACTGGTCAG AACAGATACA	6720
GAACATAAGG CCATGGCTAG ACCTGCCAAT TCTGGGTTGA GAGCCAGTCC AACACCTGAA	6780
AAGACTCCTG CTGCAATCGG AATTCCGACA ACATTGTAGA TAAAAGCCCA GAAAAGATTG	6840
AGTAGAATTC GATGAAAGGT TTTCTTACTC ATATCAAAGG CACGAACCAC TCCTAAAAGA	6900

435

TTATTGGTTG TCAACACCAA ATCTGCTGAC TCGATGGCGA TATCTGTTCC AGCTCCCAT	6960
GCAATCCCCA CATCTGCTAC ACTAAGGGCA GGAGCGTCAT TGATACCGTC CCCAACAAAG	7020
GCTACTTTCC CTGACTGTTG CAGTTTATGG ATTTTCATGG CTTTTCTTC TGGCAAGACG	7080
CCTGCAATGA CCTCTTCAAT TCCGATTGA TCTGCAATAG CACGCGCCAC ACCAGCATTG	7140
TCTCCTGTCA GCATGACTGT TCGGAGACCA CGTTTTTTTA GCTGACTGAT GGCTAGCTTA	7200
GCATTTTCCT TAGGAATATC TTGCAAAGCA AGCAAGCCTT TGATTTCATT GTCAACAGCT	7260
AAGAACACAA CTGTCTTAGC TTCTTTTTCT AGTTCTTCTA GTTTATCTTG ATAAGTATTA	7320
GAAATATCCA TGCCATCCAG CATTTTAGCA TTTCCAAGTA AAACCTGTTT TCCATTGATT	7380
CGCCCTGAAA CACCTTTCCC GTGCAAGGAC TGAAAATTTT CAACAGTTTG AAACCAAGT	7440
CCAGCTTCAC TCGCTCGCTT AACGATAGCC TCAGCCAGTG GGTGTTGAGA AGCATCTTCC	7500
AAGGAGGCTG CCAACCCAAA CACTTCTACT TCGTCGCCGA TGACATCTGT TACCACAGGT	7560
TTCCCTCCG TCAAAGTCCC GGTCTTATCA AAGACAAGGG TTTGAACTTT CTGGATTTC	7620
TGTAAGACAG TTCCATTTTT GAGGAGAACC CCCATCTTG CACTACGTCC TGCCCCACC	7680
ATAAGGCTG TCGGTGTTGC AAGTCCCAAG GCACAAGGAC AGGCGATAAT CAAAACCGCC	7740
ACTCCGTAGA GAAGAGAGGA CACAAAGCTA GCTCCAAGCA CAACCACACT ATCCCTGAGC	7800
AAGACGAACC AAACCCAAAA GGTGATGATT CCTAAAATGA CAACTACTGG GACAAAAATC	7860
CCTGAAATCT TATCCGTCAA GTCCTGAATC GCGCACGAC TTGTCTGAGC TTTCTTCACA	7920
AAATCCACAA TCTGAGCCAA AACAGTCTCT GAGCCAATT TTTCTGCTCT AAAGACAAGC	7980
GTTCACCTAT GATTGATGGT TGAGCCAATG ACAGTATCTC CAACTGTCTT GTCCACAGGC	8040
AGACTCTCAC CTGTCACCAT GGATTCGTCA ATACTAGAGA CACCTTCTAC TACGACACCA	8100
TCAACAGCAA TCTTTTCACC GGGACGCACT CGAATCAGGT CGCCTACCTT GACTTGTTC	8160
AAAGGAACTT GGACATAACT ATCATCACTC AAGACTTCTG CGGTTTTAGC TTGCAAGTCC	8220
AGTAATTTCT CCACAGCTTG GGACGTATTT TTTCTCATTT TTTCTCATAA AACTGCTCCC	8280
AAAAGAACGA AAAAGAGGAT AAATCCAGCA CTTTCGAAGT AAACAGGGAG ACCAGCAAAG	8340
AGAGCAACTA GGCTATAGAA ATAAGCCACT AGAGTTCCCA GCGCAACCAA GGTATCCATG	8400
TTGGCATTGT GCTTTTAAA ACTGGCCCAA GCACTCTGGA TATATGGCTT ACCTGCAACT	8460
AACATAATAG GCCTTGTTGC TAGAAAGGTT CCCCAATGCA TGACTTGATG ACTAATGCTA	8520
CCTGTCAACA TCCCAATCAT GAGAATCACA AGAGGCACAG TAAAGATACT ACTAATCCAA	8580
AAACGTTGCA GGAGAGATAG AGATTTTCGA GTCTTCTCAA CGACTGTATA GCTTCCCTTT	8640

436

TGCATCTTCA TGCCACAAGA AAATTCATGT CGCCCTAATT CTTGAGGCGT AAAACGAATG	8700
ACTTTCTCCT CATCTACGCC GATTGGTTCC AAGATACCTT CTTCTTCAAA CAGAAATTCC	8760
TTATAACAGT TTGAAGGAGT AGCACGATGA AAGGTAATCT CAGCTGGAAT TCCCTTTTGA	8820
AGCTGGATAT GGGCTGGATG ATAGCCTTTT TCAGCTCGGA TACGGATTTT TTGAATGCCA	8880
TTTTCTAAGC TTGCTTTCAC AATTTCTGTC ATAGTCTCCA CCTACTCTAC AATCATCTTG	8940
CCGTGCATCA TGTTCATACC ACAAGCAAAG CCAAACTCTC CAGCCTGTTC AGGCGTGATT	9000
TCCACTACAT ACTCTTCCCC CATTTGGCAGG TTCGCATGTA CACCAAAATC TGGAAAAACA	9060
ATTTGATCCA GACATGGTGA AGGATCCTTG CGGTCAAAGA CAATGCGTGC TGGCACTGAT	9120
TTCTTGAGGA CAATCAACTC AGGAGTATAG CCTCCCATGA CTTCCACTCG AATCTCTTGG	9180
TATCCGTTTT TTTGCTGGGC TTTTGTGCGA GATTTTTTCTAG GCTTTTTGAA AAACCAAAAC	9240
AAGATAAACG CGATAAGGGC AATACAAATA ATGGTTACAA TACTATTAA CATGACGTCT	9300
CCTTTACATA CAATTACATC TTAATTCTGT TACAGCACTT GATTTCTTCT CTGAAATCAC	9360
AGCTTCCAAG TCTTCCAAGT CAGTCTGAGT AAATTCACAT TCTACAATCA AGTCAGCCAA	9420
CAAAATCCTA ATCCTACGGG AACAAACCTT GTCTTTGATA TCTTGGACAA GTAAATCCCG	9480
ACTTTGGTCT AGAGTTAAAA GGGCTGAATA AACAAAGGAC TTGCCTTCTT TTTTCCGAGT	9540
CAAACACTCT TTATCAACCA GACGAGCCAA AAGTGTCTGA ACCGTGGACT TGGACCAGTC	9600
AAACCGCTCT GCCAAAACCC TAATCAAATC TGTACTGGTC TGCTCCCCCT GCATCCAAAT	9660
AATCTTCATG ACCTGCCATT CTGCATCTGA AATCTGCATT ACCATACCTC CAAAATCTAC	9720
ATTTGTCAAT TACTCATC AGTATACTCT TAAATCTAC ATTTGTCAAT TATAGAAATA	9780
ATATTTTCTT CGAAAAATAG AATTTTAATC ATTTGAAAAA CGATTTGCAG TCAAATATTA	9840
CTATATAAAC AATAAAAAATA TGCTATACTA AAGAAAAAAG AAAACAACCA CTAGGGGTGC	9900
GTAAAGCTGA GATTAACGAC TGTTAGATCC CTCTGACTCA ATCTAGGTAA TGCTAGCTGA	9960
TGGAAGTGA AATGATAATG GGGACTAGCA GTCTTCTATT GCCTTTCTAA AACAGACTAG	10020
CTTGTTCTTA AGAATACAAA CTTCACTTGG TTGGGAGGTT TTAGATGACT TATTTACCCG	10080
TTGCTTTGAC CATTGCAGGG ACTGACCCTA GTGGTGGTGC TGGCATTATG GCAGATTAA	10140
AGTCATTCCA AGCGAGAGAT GTCTATGGAA TGGCTGTTGT AACCAGTCTT GTCGCTCAAA	10200
ATACCAGAGG TGTTCACTA ATCGAGCAGC TTTCTCCTCA AATGTTGAAA GCCCAATTGG	10260
AGAGTGCTT TTCTGATATT CCACCTCAGG CTGTAAAAAC TGGAAATGTTG GCTACTACTG	10320
AAATCATGGA AATCATCCAA CCTATCTTA AAAAAGTGA TTGTCCCTAT GTCCTTGATC	10380
CTGTTATGGT TGCTACAAGT GGAGATGCCT TGATTGACTC AAATGCTAGA GACTATCTCA	10440

437

AAACAACTT ACTACCTCTA GCAACTATTA TTACGCCAAA TCTTCCTGAA GCAGAAGAGA	10500
TTGTTGGTTT TTCAATCCAT GACCCCGAAG ACATGCAGCG TGCTGGTCGC CTGATTTTAA	10560
AAGAATTTGG TCCTCAGTCT GTGGTTATCA AAGGCGGACA TCTCAAAGGT GGTGCTAAAG	10620
ATTTCTCTT TACCAAGAAT GAACAATTG TCTGGGAAAG CCCACGAATT CAAACCTGTC	10680
ACACCCATGG TACTGGATGT ACCTTTGCTG CAGTGATTAC TGCTGAACTA GCCAAGGGCA	10740
AGAGTCTTTA CCAGGCAGTT GATAAGGCCA AGGCCTTTAT CACAAAAGCT ATTCAAGATG	10800
CCCCTCAACT CGGTCAATGG TCTGGTCCAG TCAACCATAC AACTTTTAAA GATTAAGAAA	10860
AAAACTCTC TAGTTCCAC TTTAAGGGAA TTAGAGAGTT TTTATACTCT TCGAAAATCT	10920
CTTCAAACCTA CGTCAGCTTC CATCTGCAGC CTCAAACAC TGTTTTGAGC TGACTTCGTC	10980
AGTCTTATCT AAAACCTCAA GGCAGTACTT TGAGCAACCT GCGACTAGCT TTCTAGTTTA	11040
CTCTTTGATT TTCATTGAGT ATTAATTAGG AAAGAATGTT ATGCAACTTT TTTAAAAAGG	11100
CTTGCGTTTT TGCTCAATA TCTTCTGCTT GCATCAAATC ACGTACAACA GCTACACCAG	11160
CTATGCCAGT GCCCATAAGC TGATCAATAT TCTCCGAAGT CAAGCCTCCA ATAGCAACTA	11220
CTGGAATGGC AACCCTTTGG CAAATTGTTT TCAAGGTCGA TATCAGAGTA ATGGGCGCAT	11280
TTTCCTTGGT GGTGGTTGGG AAAATGGCTC CTGTACCCAA GTAATCTGCA CCTGATTCTC	11340
CCGCTTCCAG AGCTCTTTTA ACCGTTTTAG CGGTGACACC GAGGATTTTT TCAGGACCCA	11400
AGACTTTGCG AGCTACCGAA ACTGGTAATT CATCATCTCC GATATGCAGA CCTGCTGCAT	11460
CAACCGCAAG ACAAACATCC AACCAGTCAT CGATTATCAA GGTACCTGA TAAGCATCTC	11520
TTATTTCTT GACTTGTTTT GCCAGTTGAT AATATTGATT GGTGTGAGA TTTTTTCTC	11580
GCAATTGGAC TATGGTAACC CCTGAACGGC AGGCCGTCTC AACTTTTGCA AGAAAGCTTT	11640
CCACGGAATC TTGATAGCGA TTGGTTACCA GATATAGTCT AAGTGCTTCT CTATTCATAA	11700
ACCTCTCCTT TGATGGTATC TAGCCAATTT TCATCTCTTC TTAGGAGCGA AAGCTGATTG	11760
AGTACTTGGT AACGAAATTC TTCCAATCCC ATTCTTGAA CAACTATTTT CTCAGCAGCG	11820
ATATTGAGAT AAGAGACTGC TAAGCAAGAA GCTTCAAAAC CAGTCTTTCC TTGGCTGAGA	11880
AAAACAGCTG TTAAGGCTCC AACCAAGTCT CCTGTCCCTG TTATCCAGTC TAATTCAGTA	11940
CAGCCATTTT CCAGTACAGC GACCTGATTT TTCGAAACGA CGAGGTCCTT GGGACCTGTG	12000
ACTAAGAAAG ACATACCAGG ATAGGTCTGA CACCAGTCTT TCAAGACTTG AAGCAAATCC	12060
TCCGTTTCTT GATCTTTAGC ACTCGCATCG ACCCCAACGC CGTGGTGCTT TAATCCAACA	12120
AGACTTCGAA TTTCTGACAT GTTTCCTTTA AGGACCGTAG GTCTATAGTC TAAAAGGTCT	12180

438

TTAACTAAGC	TCTTACGAAT	GGATGAAGTC	GTTACGCCAA	CCGCATCTAC	TACCATCGGG	12240
AGAGAAGATT	GGTTTGCATA	CGAAGCTGCC	ATGCCGATTG	CTTTTTCCTT	CTCAGCTGAC	12300
AAATGCCCCA	AATTGATGAA	GAGAGCCTGA	CTTTGCTTAG	TAAAATCAAG	AACTTCACGG	12360
GAATCATCTG	CCATGACAGG	TTTGCATCCC	AGAGCCAAAA	TCCCATTGTC	CAGCATCTCA	12420
CAAGAAATCT	CATTGGTAAT	GCAGTGAATG	AGGGAAGTAG	AGCCTATAGG	AAAGGGATT	12480
GTAAATTCCT	GCATCAGTCT	ATCCTTTCAC	TAAAGAAATA	TCCCTGCACT	TTTTTAAAGA	12540
ATTCTGCTT	GATTAAAAAT	CGAAAGGCAA	TAAAGGAAAT	CGCTGTACCA	ATCAAGGTTG	12600
CTCCGAAAAA	TCGAGGCGTG	TAGATAAACC	AGCTAAGCTT	AGCAGCTGAT	CCTGTAAAGA	12660
GTACCATAAC	AGGATAGGAA	ACAATGGAAC	CAATAATACC	TGTTCCCAAA	ATCTCTCCTA	12720
GAGCAGAATA	GTGAAATTTT	CGACCGTACT	TATAAAAGAG	ACCTGCTAGA	AGGGCTCCAA	12780
AAGTCGCTCC	TGTGAGAGCT	AAAGGCGGAA	TCCCTTGAGT	CGTCATACGG	ATAAAGGCTG	12840
TGACTGTAGC	CATAGCCAAG	GCATAAACAG	GTCCCATCAT	GATTCTGCT	AGAATATTGA	12900
CTACACTGGA	CATCGGTGCC	ATTCCCTCAA	TTCGAAAGAT	AGGTGTAAGG	ACTACATCAA	12960
GGGCAATCAT	CATAGATAAA	ATGGTTAATT	TGTGAACTTG	TAATTGGTGC	TTTCTCATGC	13020
TTCTATTCTT	CTCCTTTTTC	TAAAGACTGT	AAATCGCTCT	TCCATGTCTG	GTGTTGGTAG	13080
GCCATTTCCC	AAAACCTGGC	TTCCATATGA	ACACTGATGT	GGAAGGCATC	TAGCATTTTT	13140
TGCTTGCTG	TCTCGTCACT	TTCTCGATAG	AGCTGATTGA	CCAGTGCTCC	CTCCTCTCTG	13200
ATCTGTTGCT	CTAACTCATC	CGTAATATAA	GTTTCAATCC	ATTGTTGATA	GAGAGGATTT	13260
GGTGATGGT	TAAGATTAAAG	TGATTTGCCT	ATATCATGGT	ATAACCAAGG	ACAAGGAAGC	13320
AAGCTTGCAA	AAGCGATGGC	TAAGTTCCGT	TCTGCAAATT	GCCTATAAA	ATGAGAAATG	13380
TAATGATAAC	AGGTTGGAGC	GATTGGATGT	TGCTCCATTT	CCTGGTCGCT	GATTTCCAAT	13440
TCCTTGAAAA	ATTGTTGGCG	AATAAATAAC	TCACCTCCA	CTAAACCCTG	AGCATTTTGT	13500
TTCAAGAGTC	TTTTCATCTC	TTGGTTTGAA	GTCTTATCAG	CCAAAAGATG	ATAGATTTCT	13560
GAGAAAGCCT	TCAGATAGTA	GGCATCCTGA	ATCAGGTAAT	AGCGGAAAA	GGCAGGTTCT	13620
AAATTCCCCCT	CTTGTAATTG	TAAAATAAAG	GGATGATGAA	AGGAAGCCTG	CCAAGCTTTC	13680
TTGGATAATT	CCATCGCAAT	ATCTGTAAAT	TCCATAATAA	CTCCTTTATA	AAAATAGACT	13740
GGTTTGAAGC	AATAAAAAGA	AAAGCAGGTA	GATTAATTTT	GTTTTTTTAG	GAATATAAAA	13800
AGTCCGATAG	CTATTCTTCA	ACTGTGCATG	TTCTGCATAT	CCGTGAGCAG	ATAGAGCTCT	13860
CAGGTAAAGA	TGGCGCCACC	TAAAGACTGT	CATCAGAACC	TTACTGTAAA	TCAAGGGCGA	13920
CCAAAAATGT	AGTTCTTGAC	CACGTAATAG	GCAAGCTTCT	TTGAGGGACT	TGATTTCTTC	13980

CTGAATGAGA GGAAAAGAAT TGAATACCAC AATCAAGGCA TAGGACCAAG AGCGTGATAG	14040
CCCCTTTTGA GCCAAGTACA AGAGAAGCTC TTTTAGTGAA ACAGAGGAAA CAAAGACAAG	14100
GCCGATACAA ACTGTCACAA AGGCCCTCGT TCCAAGCATG ACTGCCTGTG AAGCATCTCC	14160
GTGTAAC TGA ACTGCCCAGT AGTTGGCAAA AGATGGTAAA ATGGCAAGTA TGATCATCCA	14220
AGCTAACATT TTAAATCGAC GGTAATAGAG CATAAAGAGA ATACAAAATG CGACTACCGA	14280
AAGAGTCAGA GCAATCGAAG GAATGAAAGA TGTTTCCAAG GATAAAATCA GCAAGAAGAG	14340
ACTGATAATC GGTGTCTGGG TTGCTACTTT GACCATACTA TCTCACCTCC CCTTGGGTAT	14400
TGCTACTCTG AGATGTAAGT GGTTTGGTAA TGGTCACTTC TTTCACATGC CGAAGACCC	14460
GACTAGTCAT CTCAATCCAA TAATCAACCA CAGAAATCAA AGGGTCTAAA CGATGACTAA	14520
TGAGCAGAAA ACTTCTTCCT TGATTCCTCT CCTCCACAA	14580
AGGCTCTATC ATCCAAACCT GCAAAAGGTT CATCTAGCAA GATCACGGAA GCCTTACTGG	14640
TCAAGATGGT CAGGAGCTGA AGAATTTTTT GCTGACCACC ACTTAATTGA TAGGGACTCT	14700
TATCGACTGC CTGCTCCAAA TCAAAATATC GTAAAGCTTG AAAAATCCGC TGATTTCTTT	14760
CAGAAATCAGG TCCATCTAAT TGAAGCTCCT CTCCGAGACT GACTCCGATA AACTGCTTCT	14820
CAGCTTCCTG AACAAACACCA GTCAGATCAC GATACAAACT CTTTTCTTTT TTCAGGACCG	14880
AACCCCTTCCA AGTAATGCTC CCCTTATACT TTTGAAATTG AAGAATAGAC CGAAAGAGGG	14940
TTGATTTCCC GACACCATTG TCACCCAGGA TACAGGAAAT CCCTTGATAG AATGTGAAAT	15000
CAGCAATTGA AAAGAGGGGG CGATTACCAA GCTCACCAGT CACACGGTTC ATATGGAATA	15060
GTTCCGGGCT AGAAGCAACT TCCTTTGAAG CAACCTGTGT CATCTCATAG GAAGGGATT	15120
GAAACACTTC CCTTAGTTTT CCGTCTCTTA GCTCCACCAT ATGGTCCGATA TAGGCTTTAT	15180
AGTCAGATAA ATCATGGTCG CACAAAATAA CTGTCTTCCC ATCATAGACC AACTCTTTTA	15240
GAATCTCCAA TATCTCGATT CTGCTCTTGC GGTCAATGGA AGCGAAGGGC TCATCCAAGA	15300
GATAGACCCT AGGATTCATG GCAAAGAGGA CAGCCAGCGC TGCTTTTTCG TTTTCCCCAC	15360
CTGATAAGTG ATGGATGAGA CGGTGCAAGA TGTCCTTGCA ACGACATTGC TGGACAACCT	15420
CTGCTATTTT AGAATCAATT TCCTGAAGGT GATAGCCGAT ATTTTCCATG GTAAAAACCA	15480
ACTCCTCAAA CAAGCTCTCC ATGGTAAATT GATGATTAGG ATTTTGCAAG AGAATACCAA	15540
CCGTCTGGAC ACGTTCGACG ATAGAAAGCT GACTGACCTC GCTCCCATCT ATCAGGACTT	15600
GACCGCTATA GGGAAAGAGAA CTAACCTGGG CAATCATTTG AAAGAGGCTG GATTTTCCAG	15660
ACCCACTACT CCCAACTAAC AAGGTAAAGG CTTGCGCATG AAAAGTAAAA TCAAACGGCT	15720

440

CAGAGAAGAT TGGGGACTGA ATCGCTCGTA GTTCCAGACC CATCTATGCT TTTCCTCCAG	15780
TTGCCAACTG ATGATAGAGT TTGACAATGG CACGAACCAA GATGGTACAG AAGAAATAAA	15840
CAGAAATAAA ACGTACCACA AGCAAGGAAA GGACAAACGG AAGGGAAAAG GCGTAGTAAC	15900
CTAACTTAAT GTATTCATAG ACAAAGCTAA CAAGCGTAAT CCCAATACTA TTAGCAGTTA	15960
GAGAGAGCCA ACTTTCATAG CGATTCTTAG TTACGATAAA ACCAAATTCA CTCCCAAAC	16020
CTTGAACAAA GCCAGACAAA AGAGCTCCTA GACCAAATTG GCTACCATAA AGGACTTCAG	16080
CAAGCGCAGC TAGCACTTCT CCAATCGTTG CACTTCCGAC TCTCGGAACA AAGATGGCAG	16140
CAATGGGCGC AGCCATACAC CAGAGACCGA AGAGGATTTT ATTGGCAAAG GCCTGCAAAC	16200
CAAGAGGTGT TAAGAGTAGA CTGAGAATAT TATACACATA TCCTGAACCA ACGAAAACCC	16260
CACCAAAAAA GATAGACAAG AAAGCAAGCA AGATAACATC TTTTAACTGC CATTTTTTCA	16320
ACATAAAAAA CTCCTTTTTT TAAAGAAAAG TGAGGCACTC AAGAAGACCG ACCTAAATAC	16380
TTTGATAGC AGACTGAATT TAGAACAGTA CACAAGAACA CTAAATATT TCTAGAAATT	16440
AATTTGAATT TTCTAATTGA TTTGTTCCGA TCTTATTTCA ATCTACTATA TCATCTTCAT	16500
CCAGTTTCGT AAAAGAAAAA ACTCTAATTA CAGATACAAA TTAGAGTTCA GCTTACAAGA	16560
TTAGACAGTT CTTTTCGACA TACGAAAAAA ACATTTTACA TTTCCCTTCG CCAGTCTTAA	16620
CTGTATCAGG TTCAATGGGT ATCATCTCAG CCTAAAGCAC CCCAAATGTC TTTATTATTT	16680
AATTATGTGA TTATTATAAC ACACATTTTA TACTAGTTCA AGAAATTGAA CTGGAAATAC	16740
AGCCTTGAC TCACAAAGAC AGCAGATCTT TCTTTTGCAA AAAACAAATG ACCTGTTTGA	16800
TGAATTAGCC ATTCAAGCTG AATCTGGACA TAGCTTTTAA AAAAAGGAAA ATCTACTTAA	16860
CTTAGAATCC AAGGATAGAT ATCTATTGTT CACTCATTTT CCGAACAGTT TTTTCTATAT	16920
TTTTTGATA CGATATTGCC GAAATGATTG AAACGCCATC CATATTGGTC TTTATAATGT	16980
CTTTAATATG TTTGCTCTGT ATCCCACCAA TTGCAACTAA AGGCATTTGT GGCAATAGTT	17040
TTCTCATCAA TTCAAGACCT TCATAACCTA TAGTACCACC AGCATCATCC TTTGACTGGG	17100
TACCAAATAC AGGCCCAACA CCTACATAAT CTACATATTC AACTTTTGAT TGTTGAAATT	17160
CTTCTTCGTT TCTTATAGAA AGACCAATTA TTTTATCTGG CATCAATTTT CTAATTTTCA	17220
CAACACCAAT ATCATCTTGA CCTACATGTA CGCCATCGGC GTCAATTTC ATTGCTAAAT	17280
CTATATCGTC ATTAACGATA AATGGAACAT TGTATTTTTT ACAAAGTTCT TTAATTTGGA	17340
TAGCTAGCTC AAGTTTTTCT AAGCCTTCTA AAGCACCTC ACCTTTTTCT CGAAATTGAA	17400
ATAAGGTTAT ACCACCTTTT AAGGCTTCCT CAAGCACTGT ATATAGATTT TTTCCTTGGC	17460
AAGTAGTCGT TCCACAAATA AAATATAGTT TTAGTAATTC TTTATGAAAC ATCTTACTTC	17520

441

ACTCTTTTGA ATTCTTTTAC ATCTTCATCT GTAATCTCGT ATAAGGCATT TATAAATTCA	17580
ACTTTAAATG TCCCAGGAAG ATGTCCATTT GGACGTTTTT CTGCTATTTT TCCAGCGATA	17640
TTGTAAACCA ACACTGCTGT TTTTAATGAT TTCAATTCCT GACCTTTTTT TAGTCCGATA	17700
AAGCTTGCTA CTACAGCTCC TAATAAGCAT CCTGTCCCAA TGACTTTCGG CATCATAGCA	17760
CTACCATTAT GAATCATTAC CACTTCTCCA TTAACAGCAA TGGCATCCAC TTCACCTGTT	17820
ACTACTATTG GAATATTGAA CTTCTCATT TCTGCTAGAG CAATTTCTGC AATATTATCT	17880
ACGCCCGCAC TATCTACTCC TTTAGATGCC ACATCTATTC CTACTAAAGA GGCAATCTCG	17940
CCAGCATTTT CTCTAATCGC TGCTAGTTTA TAATTGTTGA TTAGATCATC TGCTACTTTT	18000
TTTCTATATT CTCTGCTCC ACAGGCTACA GGATCTAAAA CTGCTGGGAC ATTATATTTT	18060
TCTGCAATTT TCAGAGCAGC TTGGTATAAT TTCCAATTTT CATCTGTCAA TGTTCTATG	18120
TTTATTAATA AACCACCAGC ATACTTTAAC AAATCCTCTA AATCTGCTGG AAATCACTC	18180
ATGGCTGGTG AGGCGCCCAG TGCTACTAAT CCATTTGCTG TGAAATTTTT TACTACATCA	18240
TTGGTTATAC AAATGACCAA TGGTGCTTTT TCTTTTAATA ATTTAAACT TGTCATATTG	18300
AAATCCTTCC TTTTCACTTT ATACGATCTA CTAATTTTGA TTTATCTTTA GTTGAGAATT	18360
TTTTTCATTT ACATTGAATG ATTTATACTC AATGAAAATC AAAGAGCAAA CTAGGAGGCT	18420
AACCGCAGGT TGCTCAAAAC ACTGTTTTGA GGTGTGGAT AGAACTGACG TGGTTTGAAG	18480
AGATTTTCTGA AGAGTCTTAC CTCATCAAAAT TTGTAAATAT CATGAGCCTT CTCTAGACAT	18540
CGTAACCAAT ATCAAAAAAA GCTAATTCTA AAGCGACTGC TTGATTCCAG CCTTCTGAA	18600
GTCTGTCAA ATCTTCTCGA TTTTACCAG CACGATTGAG TTCGTCAACC AGAAATTGAA	18660
CCCACTCTGC AAAGAAAGGA CCTCTGTGGA GATTGATCCA TTCCGAATGA ATATAGACTT	18720
CAGGTAAAGC CAAATCTTTA GAACCCAGT CTAAATAGAG ACCTTCTGCA ATGACCAGCA	18780
TGACCAAAAG ATGGGCATAG TCTGATGAAG CCACCGCCGA ATACATTAGA TCCTGAAAGC	18840
CTTTTGTTAC AGGGTGCAA GTCACTTCTA GATAGTCATT CTCTGCTACT TTAACTCTT	18900
TAAAAGCCTT TTGGAATAA CCATCTTCAT CTGCTTCAAG AAAGCCTAGT TGCTTGGCAA	18960
AACGAAGCTT GGATTCAAGT TTATCTGCGT GACTACGCAG GCACCCAGCA TGGATAAGAA	19020
GGCATCAAAG AAGTGATAAT CTTGAATCAG ATAGTCCTTT AAGACCTTAT TCTCAATTGT	19080
CCCCGAAAA AGTTCCTTAA CAAAACGATG ATTGATTGCA GCCTGCCAAT CTTTCTGACT	19140
GCTTTTTAAT AATTCTCCAA CAGTCAAACC TGGCTGAAAT GCATAGCTT GTGTTTCCAT	19200
ATTACTTCT CCTCTCTTTA CTGTGTTAGTA ATTAATAAAA CACCAAGAAA TATCAAGCAA	19260



442

AATCGTAATT CCACTTGATC CTTTAAAGC ACATCGAGAG CATTTCAGAGAG GAGCTAACTA	19320
AACAAGCCTA TCCAGTTTAT ATAAACAAAA AACTCCAATT ACAATCAAGA ATTAGAGTTG	19380
ACTTACAAGA TTAGACCGTT CATTTCACCA TACGAAAAAA CTGTTCCACAT TTCCCTTCGC	19440
CAGTCTTAAC TGTATCAGGT TCAATGGGTA TTATCTCAGC CTAAAGCACC CCAAATGTCT	19500
CTATTATTTA ACTACTGAAC CAGTATAGCA AAAAATGAAA GCCCTAGCAA GATATTTGAC	19560
CGAAAAATAT CTTTATATAT AATATATTGA AACTAGAATA GTACACCTCT ACTTATAAAA	19620
CATTGTTAGA AATCGATTG ACTGTCCTGA TTGATTGTGCT CTATTCTTAT TTCATTTTAC	19680
TATAGTTTTC GATAGCAATT TATTCTTCCA ATACACGAAG AAAAACCTCC ACATTTCAGTG	19740
GAGGCAATCT GTTTTATCAA TACAATTTTA AGTCACGAGG GTCAACTGGG AAGGTTGGGT	19800
TGTATGGATT GTGACGGAGC TTGAAGTGTT TGACATCTTC AATGGTCTGA GTTCCAGACA	19860
ATTGCATAAC TGTCTTCAAT TCCGCATTCA AGTGTTCAAA GACTTGACGC ACACCGACAC	19920
TACCACCGAG AGCCAAGCCA TAGATGACAG GCGGTCCAAT AGCAACCAAG TCTGCTCCTG	19980
ATGCCAAGGC TTAAAGACG TGTGACCAC GACGAACACC AGAGTCAAAG ACAATCGGCA	20040
CACGTCTATC AACTGCTTCT GCCACTTCTT GAAGCGAGTC AAAGGCAGCT GGTCCACCGT	20100
CGATTTGACG ACCACCGTGG TTGGTTACCC AGATACCAGA AGCTCCTGCA GCAAGCGAAC	20160
GTTCAACGTC CTCACGGCAT TGTGGTCCCT TGACATACAC AGGAAGACCA GAGTATTCAG	20220
CGATAAATTC TACATCGCGT GGAGACAAGC GTTGTTTAGC TGATTGTAA ACAAAAGTCCA	20280
TTGATTTACC AGCACCTTCT GGCAGGTATT CTTCAACAAT CGGCATGCCA ACTGGGAAGA	20340
CAAAACCATT ACGCTTATCC ACTTCACGAT TCCCCCTAC AGTAGCATCT GCCGTCAAGA	20400
CAATCGCTTT ATAACCTTCA GCCTTCACAC GGTCCATGAT GTGGCGGTTG ATACCGTCAT	20460
CCTTACTAAA GTAAAATTGA AACCAATGAG GTGTCCCTTG GAGGGCTTCA GAAATCTCTG	20520
GAAGGTCAAC AGTAGAGTAA GAACTGGTTG TATAAAGAGA ACCAAACTCA TGCACACCAC	20580
GCGCAGTCGC CACTTCCCC TGTTCATTG CCAATTTATG AGCCGCAACA GGTGCCATAA	20640
TGATTGGAGA AGATAGTTTT TCACCTGCAA ATTCAATCTC TGTACTTGGA TTTTCTACAT	20700
TGCAAAGTGT ATGAGGAACG ATGAGCTTGT GGTAAAGGC ACGGATATTC TCTCTTAAAG	20760
TGAAAGTATC TTCCGCCCCA CTAGCGATAT AGCCAAATGC TGCTTTAGGA ATAACTTGTT	20820
GCGCCATTGG CTCCAAATCA TAGGTATTGA TGAATCTAC ATGACCTTCT GCATTGCTTG	20880
TTTTGTATGA CATAAAATGT CCTCCTTAAT AAGTAAGCGT TTACTTTGTG TATTACAAAA	20940
ATATCTTAAC TCTTTTCAA AACTTTTAAA ATATTTTCTT TGGAAATTC AGAAATTTTA	21000
TGTCTATGAT AAAAACTCTT ATAACGGCAA TAAAAATAG ATATTATCCA AAGAAGATT	21060

443

TAAGTGCTAC AATAACTGTA TTATTTCTAG ATGGGAGGTT CTATTTTGG ATTGATCCAT	21120
TGTTGAACAA TATCTACCAC TATATCAAAA GGCATTCTTT CTGACCTTGC ATATTGCAGT	21180
TTGGGGAATT TTGGGATCCT TTCTGCTCGG TTTAATCGTT AGTATCATCC GACATTATCG	21240
AATCCTTGTT TTGGCGCAAG TAGCGACAGC CTACATTGAA TTGTCACGTA ATACGCCCCT	21300
TTTGATTCAA CTCTTCTTC TCTACTTCGG TCTTCCCCGA ATCGGGATTG TCCTATCTTC	21360
AGAAGTCTGT GCAACGCTTG GGCTTGTCTT TTTAGGAGGC TCCTATATGG CAGAATCTTT	21420
CCGAAGTGGG CTGGAAGCCA TCAGTCAAAC CCAGCAGGAG ATTGGCCTCG CTATTGGTCT	21480
GACACCTCTA CAGGTCTTTT ACTATGTGGT TCTTCCGCAA GCAACAGCGG TGGCACTCCC	21540
CTCCTTTAGT GCCAATGTCA TTTTCTTAT CAAGGAAACC TCTGTTTTCT CAGCAGTGGC	21600
TTTGGCCGAC CTCATGTACG TCGCCAAGGA TTTGATTGGT CTCTACTATG AGACAGACAT	21660
TGCGCTAGCT ATGTTGGTAG TTGCTTATCT AATCATGCTG CTACCCATCT CACTGGTCTT	21720
TAGCTGGATA GAAAGGAGGC TCCGCCATGC AGGATTCGGG AATCCAAGTA CTCTTTCAAG	21780
GAAATAATCT CCTGAGAATC TTACAGGGAT TGGGCGTTAC GATTGGGATA TCCATCCTGT	21840
CTGTCCTCTT ATCCATGATG TTCAGAACAG TCATGGGAAT CATCATGACC TCCCATTCTA	21900
GAATCATACG ATTTTAAACA CGATTGTATC TGGAAATTTAT CCGTATCATG CCCCAGCTGG	21960
TGCTACTCTT CATCGTTTAC TTTGGCTTGG CTCGAAACTT TAATATCAAT ATCTCAGGTG	22020
AGACTTCAGC TATTATCGTT TTTACCTCTT GGGGAACAGC TGAAATGGGA GACTTGGTAC	22080
GTGGAGCTAT CACTTCTCTC CCTAAACATC AGTTTGAAAG TGGACAGGCA CTCGGCTTGA	22140
CTAATGTTCA ACTTTACTAC CACATCATCA TCCCACAAGT CTTAAGAAGA CTGCTACCGC	22200
AGGCTATCAA TCTTGCTACT CGGATGATTA AAACCACTTC ATTAGTTGTT TTGATTGGGG	22260
TTGTGGAAGT GACCAAAGTT GGACAACAAA TCATCGATAG CAATCGCCTG ACCATCCCAA	22320
CTGCTTCATT TTGGATTTAT GGAACCATTC TAATCTTATA TTTCGCAGTT TGCTACCCTA	22380
TTTCCAAACT ATCCACTCAC TTAGAAAAAC ATTGGAGAAA CTTAAATGTCT GAAACTATCT	22440
TAGAAATCAA GGAACATAAA AAATCCTTCG GAGACAATCC CATCCTCCAA GGAATTTCTC	22500
TAGAAATCAA AAAAGGGGAA GTTGTGTGCA TCCTAGGGCC ATCTGGTTGT GGGAAAAGTA	22560
CCCTCCTTCG TTGCCTCAAC GGCTTAGAAA GTATTCAAGG TGGAGATATT CTTCTGGATG	22620
GTCAGTCTAT COTTGAAAAT AAAAAAGATT TTCACCTAGT TCGCCAAAAG ATTGGCATGG	22680
TCTTTCAAAG TTATGAACTC TTTCCCCATC TGGATGTCTT ACAAACCTC ATCCTAGGCC	22740
CTATCAAAGC TCAAGGAAGG GACAAGAAAG AAGTAACGGA AGAAGCTTTG CAATTACTAG	22800

444

AGCGTGTCCG TTTGCTGGAT AAACAACATA GCTTTGCCCG TCAATTATCT GGTGGACAGA	22960
AGCAACGTGT TGCAATTGTC CGTGCCCTCC TAATGCATCC AGAAATCATC CTTTTTGACG	22920
AGGTGACTGC TTCGCTGGAT CCAGAAATGG TGCCTGAGGT GCTGGAACCT ATCAATGATT	22980
TGGCCCAAGA AGGCCGTACC ATGATTTTAG TAACCCACGA AATGCAGTTT GCCCAAGCCA	23040
TTACTGACCG GATTATCTTC CTCGACCAAG GGAAAATCGC TGAAGAAGGA ACAGCTCAAG	23100
CCTTCTTTAC CAATCCGCAA ACCAAACGAG CCCAGGAATT TTTAAACGTC TTTGACTTTA	23160
GCCAATTCCG CTCATATCTA TAAAGGAGAT TCTTATGAAA CTATTCAAAC CACTCTTAAC	23220
TGTTTTAGCA CTTGCCTTTG CCCTTATCTT TATCACTGCT TGTAGCTCAG GTGGAACGC	23280
TGGTTCATCC TCTGGAAAA CAACTGCCAA AGCTCGCACT ATCGATGAAA TCAAAAAAAG	23340
CGGTGAACCT CGAATCGCCG TGTTTGGAGA TAAAAAACCG TTTGGCTACG TTGACAATGA	23400
TGGTTCTTAC CAAGGCTACG CTACGATATT GAACTAGGGA ACCAACTAGC TCAAGACCTT	23460
GCTGTCAAGG TTAAATACAT TTCAGTCGAT GCTGCCAACC GTGCGGAATA CTTGATTTCA	23520
AACAAGGTAG ATATTACTCT TGCTAACTTT ACAGTAACTG ACGAACGTAA GAAACAAGTT	23580
GATTTTGCCC TTCCATATAT GAAAGTTTCT CTGGGTGTCG TATCACCTAA GACTGGTCTC	23640
ATTACAGACG TCAAACAAC TGAAGGTAAA ACCTTAATTG TCACAAAAGG AACGACTGCT	23700
GAGACTTATT TTGAAAAGAA TCATCCAGAA ATCAAACCTC AAAAATACGA CCAATACAGT	23760
GACTCTTACC AAGCTCTCT TGACGGACGT GGAGATGCCT TTTCAACTGA CAATACGGAA	23820
GTTCTAGCTT GGGCGCTTGA AAATAAAGGA TTTGAAGTAG GAATTACTTC CCTCGGTGAT	23880
CCCATACCA TTGCCGCAGC AGTTCAAAAA GCAACCAAG AATTGCTAGA CTTCTCAAT	23940
AAAGATATTG AAAAATTAGG CAAGGAAAAC TTCTTCCACA AGGCCTATGA AAAGACACTT	24000
CACCCAACCT ACGGTGACGC TGCTAAAGCA GATGACCTGG TTGTTGAAGG TGGAAAAGTT	24060
GATTAGTCAT TAACTCTTAA AAGGAACTGG ATTTTAAAGT CCAATCCCTT TTTAAGATT	24120
TACCTATAAC ATCCTGAGTC TATCTAAGAT GTTCAATCTG AACACAGTGT ACATACTTTA	24180
TCTTCTATTG CATATACTTT ATCACATAAG ATACGAATAT CCTCTTCACT ATGACTAGCA	24240
ATCAAAATTG TTGTCCCTTT TTTACTAGAG AGCTTTCTAA ACAATGTTCT CATATTTTCT	24300
ACACTTGATT TATCCAAGGC ATTCATAGGT TCATCTAGTA AAAGAATAGA GGGATTCTCC	24360
ATAATTGCTT GAGCAATCCC TAGCTTTTTC CTCATACCTA GCGAATAAGT TTTAACTTTT	24420
TGGTCTTTTT GCTCATATAG ACCAACTATT TTCAGTGTAT CATTGATTTT CTGATTACCA	24480
ACTACTCCTC GTATGCTTGC CAAATATTGT AAATTCTTAA AGCCACTATA ATAATTTATA	24540
AAACCAGGTT CTTCAATCAA AGCTCCCAAA TTAGCTGGAA TTTTCTCTC AGGAACAATA	24600

445

TTTTCCCAT TGATTAACAC TTCTCCATAA GACGGACTAT ATAAACCAGC TATTAATTTA	24660
AACAATACAC TTTTCCCTGA GCCATTTCGA CCAGTAATTC CTATAATTTT CCCCTGTTTA	24720
CAACTAAAGT TAAGGTTTTG AAAAACACAT GTCTTTTTTA ATTTCAACTC AATATTTTTT	24780
AATGTAATTA TTTTATTTCAT TCTATAAACC TCCTCTTTTG ACGAGTGAAA TAGAAAATGC	24840
TTTGAAAAAG AAAGACTAAA AATAGCAACT GAAGAAATAA ATCTCGTCCT ATATCTCCAT	24900
TCCCTCGATT CAAAATATAA AATAGATAAT TAGTTCGATT TCCTACAAAT AGACCACCAA	24960
ACACAATCAT GAGTAAAAAG AAATAACGC AAGCAAAGTT CG	25002

(2) INFORMATION FOR SEQ ID NO: 49:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11443 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

CAGGTACGGT GAGGCGCAAC TAAAATATAA TTTTCATCTT GATTAGGAAT TTTATCAGTA	60
TTATGATAGT GAGCATTGCC ATTGATGGAC CATAAGAGCA ATACAATAA TCCACGCAAA	120
TAAGTATAAA ACATGCGATC TCCTTCGATT GTTTTCTTGT TATTATTATA CCTTATCAAA	180
GGAGGGCTGG CAACTTTTTC CCTTGACTAG ATACATATTT AGGATGAAAT TAGAATTCTG	240
TTAAAAAAA TGATATAATA GAATTTATGG ATAAAAATAA GATTATGGGA TTAACCCAAA	300
GAGAAGTCAA GGAAAGACAG GCTGAGGGT TGGTCAATCA CTTTACCGCA TCAGCCAGTA	360
CCAGCACTTG GCAAAATCGTT AAACGAAATG TCTTTACCTT TTTTAACGCT TTGAACTTTG	420
CCATTGCTTT GGCTCTTGCC TTTGTGCAGG CTTGGAGCAA TCTGGTCTTC TTTGCTGTTA	480
TCTGCTTTAA CGCTTTTTCT GGGATTGTGA CCGAGCTACG AGCCAAACAC ATGGTGGACA	540
AGCTCAATCT CATGACCAAG GAAAAGGTCA AAACCATCCG TGATGGTCAG GAAGTTGCTC	600
TTAATCCTGA AGAATTAGTG CTAGGAGATG TCATTGCTTT GTCTGCAGGA GAGCAGATTC	660
CTAGTGATGC CTTGGTTTTG GAAGGCTTTG CGGAAGTCAA TGAAGCCATG TTAACGGGAG	720
AAAGTGATTT GGTGCAAAAG GAAGTTGACG GCTTACTTTT GTCAGGAAGT TTCCTAGCCA	780
GTGGGTCACT TTTATCTCAA GTTCACCATG TCGGTGCAGA CAACTATGCT GCCAAACTCA	840
TGCTTGAGGC TAAGACCGTT AAACCCATCA ACTCCCGTAT CATGAAATCG CTGGACAAGT	900
TGGCTGGTTT TACTGGGAAG ATTATCATTC CTTTGGTCTT GGCTCTCTTG CTGGAAGCCT	960

446

TGCTTTTAAA AGGCCTGCCT CTCAAGTCAT CCGTTGTAAA CTCGTCGACA GCTCTTTTGG	1020
GAATGTTGCC TAAGGGAATT GCCCTTTTGA CCATTACTTC GCTCTTGACT GCAGTGATTA	1080
AGTTGGGCTT GAAAAAGGTC TTGGTGCAGG AGATGTACTC TGTTGAGACC TTGGCGCGCG	1140
TGGATATGCT CTGTCTGGAC AAGACGGGTA CCATCACCCA AGGAAAGATG CAGGTGGAGG	1200
CTGTTCTTCC GTTGACGGAA ACGTATGGTG AAGAGGCTAT TGCCAGCATC TTGACTAGCT	1260
ACATGGCCCA TAGTGAGGAT AAGAATCCAA CTGCCCAGC CATTGCGCCAG CGTTTTGTGG	1320
GAGATGTTCC TTATCCTATG ATTTCCAATC TTCCCTTCTC GAGCGACCGC AAGTGGGGGG	1380
CTATGGAGTT ACAAGGCTTG GGGACAGTTT TCTTAGGGGC ACCTGAGATG TTGCTTGATT	1440
CTGAAGTCCC AGAAGCTAGG CAGGCCTTGG AGAGAGGATC ACGTGTCTTG GTCTTAGCTC	1500
TCAGTCAGGA GAAATTAGAC CATCACAAAC CACAGAAACC ATCTGATATT CAGGCTCTAG	1560
CCTTGCTGGA AATCTTGAC CCCATTGAG AGGGAGCAGC AGAGACGCTG GACTATCTCC	1620
GTCTCAGGA GGTGGGACTC AAGATTATCT CTGGTGACAA TCCAGTTACG GTGTCCAGCA	1680
TTGCCCAGAA GGCTGGTATT GCGGACTATC ACAGCTATGT AGATTGCTCA AAAATCACCG	1740
ATGAGGAATT GATGGCCATG GCGGAGGAGA CAGCTATTTT CGGACGTGTT TCCCTCATC	1800
AAAAGAACT CATCATCCAA ACGTTGAAAA AAGCGGGACA TACAACGGCT ATGACAGGGG	1860
ACGGGGTTAA TGATATCTTG GCCCTTCGTG AGGCGGATTG TTCTATCGTG ATGGCGGAGG	1920
GGGATCCAGC AACCCGTCAG ATTGCCAATC TGTTCTCTT GAACTCAGAC TTAAATGATG	1980
TTCTGAGAT TCTCTTCGAG GGTGTCGCG TGCTCAATAA CATTGCCAC ATCGCCCCGA	2040
TTTTCTGAT AAAGACCATC TATTCCTTCC TGTTAGCAGT CATCTGTATT GCCAGTGCTT	2100
TACTAGGTCG GTCAGAGTGG ATTTTGATTT TCCCTTCAT TCCGATCCAG ATTACCATGA	2160
TTGACCAGTT TGTGGAAGGT TTCCCACCAT TCGTTCTGAC TTTTGAGCGA AATATCAAAC	2220
CTGTTGAGCA GAATTTCTC AGAAAATCCA TGCTTCGTGC CCTACCAAGC GCTCTCATGG	2280
TCGTCTCAG CGTCCTGTTT GTGAAAATGT TTGGCGCGAG TCAAGGTTGG TCTGAGTTAG	2340
AAATCTCAAC TCTACTCTAT TATCTCTTGG GGTCAATTGG TTCTTATCC GTATTTAGAG	2400
CCTGCATGCC ATTTACCCTA TGGCGTGTCC TCTTGATTGT TTGGTCAGTA GGAGGTTTCC	2460
TAGCCACAGC TCTCTTCCCA AGAATTCAAA AACTGCTTGA AATTTCAACC TTAACAGAAC	2520
AAACGTTGCC TGTTTATGGT GTCATGATGT TGGTCTTAC CGTGATTTTC ATCCTGACCA	2580
GTGTTACCA AGCGAAAAAA TAAATCAAAA CCACCAGTGT GAACTGGTGG TTTGTTCTGC	2640
GGCTATAAGC CGCTTCTACC GGCCAGGGCC AAAGGCCAC CGAAATAGCT TCTCGCGCA	2700
CCACTTTCCC GAGCAGGTGC TAAAGCACCT TAGTTACTTC CTCTTATTTA TTTGCGCAGT	2760

447

AAACGGATCT ACTGACTCGA ATAACGTGAG CTGGTCTGCT ACTCTGTCTT CTTGTAATTG	2820
ATTCTGAATA TATTCAGCTA TCACTTTCTG ATTACGGCCT ACCGTATCTA CATAATAGCC	2880
TCTACACCAA AACTTGCGAT TGCCATATTT GTATTTTAAA TTCGCATGCT TATCAAAAAT	2940
CATCAAACCTG CTCTTGCCCT TTAATAGCC CATAAAGGAC GAAACACTAA GTTTCGGAGG	3000
AATACTGATA AGCATGTGAA TATGGTCTGA ACAAGCATTG GCTTCATGGA TTATTACACC	3060
CTTACGCTCA CATAAGTCAC GTATGATTCT TCCGATACTA GCTTTGTATC TGCCATAAAT	3120
GATTTGACGA CGATATTTGG GTGCAAAAAC AATATGATAT TTACAATTCC ATGTGGTATG	3180
TGATAAACTT TGATTATCCT CTCTCATGAG GTACCTCCTG TATGATATGT TGTAGTGGCG	3240
GAGAAACCAC TTCTATCTTA TCATTTTAGG AGGTTCTTTT TGTTACCACG CTAAGAGCTC	3300
TATGGAACCA CTAGCATAGC TAGTGGTTTT CGGGAGACAA CAAGAAAGAC TGCAATCTGT	3360
GGATTGCAGT TTTTATACG ATGGATCTAT CGTAGATCTG ATGTGCAAGG CCTACGTGCC	3420
GATCATCTAT CGGTGAACCC AAGAGCGACC CTCAAGCCTG CTTGGATTGA GGTAATAGAT	3480
TCAATATCTT GTAGTTAGAC TATTTGAAGT TTGATGTAAG AAAGAGAAAG CGACAGATTG	3540
AAGTAATTTT AACTCTCTTC TATTGCTAGA ACAAATGGTC GGATAGGTTG GTAGTTTGAA	3600
AATGAAGATG CTATCTATTG TTAATGGAA CATAGTGTTA TTTATTAGAA AATCGTTTGG	3660
TTTATTTCTT ATCAAATACG AAAAGCAACT TAAATATTTT AACTAAAATA GATGTTATGA	3720
AGAAAAGGTA AAATGATTTT GGCATAGTGA GGTCTGTTC TATTTGATAT CATATTTTGT	3780
ATAAAAACAA AAATGTCCAT TGCAAAGGAC AAAATGCGAA GTATATTATT TTTTGAAAGC	3840
GATATAATGG ATTCATAAAG GAGGTGTATC GTGTCTAGAA AACAAGAACA AATGGAAACG	3900
TTGTTGCTCC TTTTGCGAGA TAGTAAGGAT TATATATCTG CTAAAGTATT GGGAGAAAAA	3960
TTAAATTGCT CTGATAAAAC GGTTTATCGC CTTGTCAAGG GAATCAACAA AGATTGTCCG	4020
GTAGAAGCAT TCATTTTATC TGAAAAAGGC AGAGGTTTCA AATTAAATCC AAGAAGTTCC	4080
CTCGTGGACG TTGATGGGAA TTTTACAGAG GCTTTTGATC CTGAAGTAAG GCGTGAAAAA	4140
TTACTAGAAC GTCTCTTGT GACTGCTCCT AAGCCACATT CTATTTATGA TTTAGGAGAG	4200
GAATTCTACG TAAGCGAGTC AGTAGTACTA AAAGATCGTC AGATATTACA AGAGAGTCTA	4260
GCAATTTATG GGTTAGATTT AAAAATGAGA CAACGAAAGC TTTTATTGA TGGGGATGAG	4320
GCTCAAATTC GTTCAGCCAT TCTAAATCTA CTGCCAATGT TTAATCAGTT GGATTAGAG	4380
CAAATTACAC AGAATAAGGT TCAGCCTCTT GACGGAGAAC TTGCTCACTT TTGTTTGGGA	4440
TTACTGATTA CACTTGAGAG AGAATTGGGG GTAAACATTC CCTATCCATA TAATATAAAT	4500

448

ATTTTCTCTC ACCTGTATAT TTTTATCAGT AGGAATCGTC GTAGTACTAG TATTCATGTT	4560
GTAGCACCTT CAAAACCTAC TATTGTTGAT GAGAAAATTT ACAGTGTCTG TCAAAAAATT	4620
ATTCAAGAAA TTGAACAATA TTTTAGGATG AAGGTTGATG CAGTTGAGAT TGAATATCTT	4680
TATCAATACG TTGTATCTTC GAGATTGCAA AAACCATTTT CTTCCGGGAA GCTTCCTTTT	4740
TCTCAGCGAG TTTTAGATGT CACTCATTAC TATTTTAGCC GTATGTGTAT GGACAATAGA	4800
GAGATTGAAA CGACAGATCC TGACTTTGTT GACTTGGCGA GTCATATCAG TCCCTTACTG	4860
AGGAGATTAG ATAATAGAGT ACAGATTAAG AATAGTCTTT TATCACAAAT TCTTTTAACC	4920
TATCCTAATC TGGTTAAAGA GTTAACAACT ATTCTAAAG AAGTGAGTCT AGTATTTGGT	4980
TTTGCTTCCT TGAGTCTGGA CGAGATTGGT TTTCTAGTCT TATATTTTGC ACGGTTTCAA	5040
GAAAAGCGAG CACGTCCTCT AAAAAAGTA GTGATGTGTA CATCAGGTGT CGGAACCTCA	5100
GAGCTTTTAC GAGCACGATT AGAAAAGCAA TTTTCTGAAT TGGATATTAT TGATGTAGTT	5160
GCTTATCATC AATTAGATGA GCTGATAAAT CTATATCCAG ATTTAGATTT CATTGTGACG	5220
ACGGTAGCTT TGCAGGAACC AGCAAGTGTC CCGTTTGTCC TAGTTAGTGT TTTTCTAACC	5280
GAGGGTGATA AACACGCTCT TCAAGCAAAA ATTCAGGAGA TAAACTATGA ATAATCTTTC	5340
GCTTGTCCTT ATGGATATAT CTGTTCAAAA TCGTCAAGAA GCCTACAAAG AATTAGCAAA	5400
TCAAATCAGC CTTCTTGTTT CTGAAGATAC AGAAAAATA GAAGAGCTTC TATATTACCG	5460
TGAGAGACAG GGAAGTATAG AGGTTGCTAA AGGTGTTCTT CTACCACATT GTGAAGGAAA	5520
CTTTCACAT CATGTCTTAG TGATTACTAG ATTAATAATCA CCTATCAGAG AATGCTCGAA	5580
GGATATCCAG TGTGTTGACC TTATTATCGG TTTGGCCATT GCAGTATCAC AGGACAAGTC	5640
ATGTATTAAA ACATTGATGA GAAGACTAGC AGATGAATCA TTCATAAATC AATTAAAACA	5700
GTTAACAAAA GAAGAATTAC GGGAGATAAT ATATGGAAAT CAAAGATATT CTTAATGTGA	5760
GTCTGATCCA GACGGATTTA CAGATGCAGA GCAAAGAAGA GGTTTTGTAG GCATTAGCTC	5820
AACTATTGGT TGAGACGGGT TATGTGTCTG ATAGAGACCA ATTTATCGAA GGTCTTTATC	5880
AGAGAGAGGC AGAAGGACAG ACCGGTATTG GGAATTATAT TGCTATTCCC CATAGCAAGA	5940
GTTCTGCTGT GGAGAAGGCG GGGGTAGTCA TAGCTATAAA TCACAATGAG ATTCCTTGGG	6000
AGACCATTGA TGGGAAAGGG GTCAAAGTAA TTGTACTCTT TGCAGTTGGT GATGATACAG	6060
AAGCTGCTAG GGAGCATTTG AAGACCTTAT CACTCTTGC TCGAAAACCT GGTAATGACG	6120
AAGTTGTTGC CAAATTAGTT CGGGCTCAGA CATCTGATGA TGTGATTGCA GCTTTTGTGTT	6180
AATAAGAAAA AATTTTGGAG GGTATCCGTA TGAAAATTGT TGGTGTGCA GCTTGTACTG	6240
TGGGAATTGC CCACACTTAT ATTGCACAGG AAAAAATTAGA GAATGCCGCA AAGGTAGCTG	6300

449

GACATGTGAT TCATGTTGAG ACTCAGGGGA CAATAGGGGT AGAAAATGAA TTGAGTCAAG	6360
AGCAGATTGA TGCAGCGGAT GTAGTTATTT TAGCAGTTGA TGTTAAGATT TCTGGTATGG	6420
AACGCTTTGA GGGTAAAAAG ATTATCAAGG TTCCAACAGA AGTGGCAGTC AAATCTCCCA	6480
ATAAACTGAT TGCTAAAGCT GTTGAGATTG TTACGAAATA ACTGAAAATA TTTAAGGAGA	6540
AAATATATGT TGAACACCTT AAACCTAAAA GGTCACTTAT TGACAGCCAT TTCTATATG	6600
ATTCCAATTG TTTGTGCTGC AGGATTCTTA GTTGCCATTG GTTTAGCAAT GGGGGGTGGT	6660
GTTCCTGACG CTCTTGTAGC AGGAAAATTC ACTATCTGGG ATGCTTTAGC AACTATGGGT	6720
GGTAAAGCCC TTGGTCTCTT GCCAGTTGTT ATTGCTACAG GTTTGTCTTA CTCGATTGCT	6780
GGTAAGCCAG GGATTGCACC AGGTTTTGTT GTTGGTCTAA TTGCCAATTC TGTGTTTCA	6840
GGGTTTATCG GTGGTATCTT GGGAGGTTAT ATAGCTGGTT TCTTGGTTCA AGCGATTATT	6900
AAAAAGGTCA AAGTACCAA CTGGATTAAA GGTTAATGC CAACCTTGAT TATTCCTTTT	6960
GTAGCCTCTT TGGTAAGTAG TTTGATTATG ATTTATATTA TTGGAGCGCC TATCGCAGCC	7020
TTTACCAACT GGTGACGAG CTTATTACAA AGCTTGGGAA GTGCTTCAAA TGGTTTGATG	7080
GGGGCAGTTA TTGGAATTCT CAGTGCTGTT GACTTTGGTG GCCCACTTAA TAAAACAGTC	7140
TATGCGTTTG TGTGACTTTT ACAGGCTGAA GGTGTGAAAG AACCATTGAC TGCTTTACAA	7200
TTGGTGAATA CTGCTACACC AGTTGGATTT GGATTGGCCT ATTTTATCGC GAAATTACTC	7260
AAAAAAATA TCTATACTCA AGAGGAAATC GAAACATTGA AATCGGCTGT TCCTATGGGG	7320
ATTGTCAATA TTGTTGAAGG TGTAATCCG ATTGTTATGA ATAACTGGT TCCAGGTCTC	7380
ATTGCAACAG GTATCGGTGG TGCTGTTGGT GGTGCTGTTT CTTTGACAAAT GGGTGTGAT	7440
TCTGCTGTGC CATTGCTGG AGTCTTATG TTACCAACCA TGACTCGTCC ACTAGCTGGT	7500
ATTTGTGCCT TGTTAGCTAA CATTGTAGTC ACAGGACTTG TCTACCGGAT TTTGAAAAA	7560
CCAATAAAAC ATGCAGAACC AGTTATGACT GTTGAAGAAG AGATTGATTT GTCAGATATT	7620
GAAATTTTGT AAGAGGGTAA CGATGTCAAG AATTGAATTT TCACCATCTT TGATGACCAT	7680
GGATTGGAC AAATCAAAG AGCAGATTAC TTTTGTGAAT GATAAAGTAG CATCTTATCA	7740
TATCGATATT ATGGATGGCC ATTTGTTC CAATATTACC TTGTCTCCTT GGTTCAATCA	7800
AGAAGTTCAA AAAATTAGTG ACACACCTTT ATCAGTTCAT CTGATGGTCA CAGACCCAAC	7860
CTTTTGGGTA GATCAAGTTC TCGATTTACA ATGTGAGTAT ATTTGTATTG ATGCTGAAGT	7920
TCTGAATGGT CTTGCTTTTC GTTTGATTGA TAAAATTCAT GATGCAGGTC TAAAGGCTGG	7980
TGTTGTCTCTT AATCCTGAAA CACCTGTTTC TACAATCTTT CCCTACATTG ATTTACTTGA	8040



451

AGCGGATGCC CGTGAACTC AAGCGGCTTG GCATCATGCC TTGACAGTA CCACCACTCC	9900
AACTGTCATT GTCTTAACCC GTCAAACTT GGTAGTTGAA GAAGGGACAG ACTTTGGTAA	9960
GGTCGCTAAA GGAGCCTACG TCGTGTATGA TACCCCGGGA TTTGATACTA TTATCATTGC	10020
TACAGGATCT GAGGTCAATC TAGCTATCAA AGCTGCTAAG GAATTGGTTT TACAAGGTGG	10080
TAAAGTACGT GTGGTATCTA TGCCCTCAAC CGAATATTT GATGCTCAAG ATGCTACCTA	10140
CAAGGAAGAC ATTTTACCAT CTAAGACTCG TCGTCGTGTG GCCATTGAAA TGGCAGCGAC	10200
CCAAAGTTGG TACAAGTATG TTGGTTTGGA TGGCGCGTC ATCGGTATTG ACATCTTCGG	10260
TGCGTCTGCC CCAGCTCAGA CTGTGATTGA TAATTATGGA TTTACGGTAG AGAATATCGT	10320
TGCTCAAGTT AAGTCCCTAT AGAAACCAAT TACAATGAAG ATACAGCTGT TGTCAGACTA	10380
GCAGATGTAG TGATAGACAC TAATCAGATG ATTGGTTATT TAAAACTGT AATGAAAATG	10440
TAATAATTTA TCTACGAAAG TTATAGTAGA TAGTATACAC AATAGAGTAT ACCCTGAAAC	10500
GGTTGCGAAG TACGCTAATC ACTTTGCTAC TGATCTAGAT AGTTTCTTTA ATCAATAAAC	10560
ACAGCATCCA CAGATTGACT TAGGATATTG TAAGTTTTTT GAAAGCTAGA GAGAAGGTCT	10620
CTAAAATTAA AAAACGCATA GTATAGGATG TTGAAATGAT GAACTGCACC CAAAAGTTA	10680
GACAGAAAAA AATCTAACTT TTGGGGTGTT TTTATTATGA AATTAACTTA TGATGATAAA	10740
GTTCAAGTTCT ATGAACTTAG AAAACAAGGA TATATCTTAG AGAAGCTTTC AAATAAATTT	10800
GGGATAAATA ATTCTAATCT TAGGTACATG ATTAAATTGA TTGATCGTTA CGGAATAGAG	10860
TTCTGCAAAA AAGGGAAAAA TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT	10920
AAAGTCTGAC ATGAAGCCTG GACTAAAGAT AGAGTTTCTC TTGAATACGG TCTCCCAAGT	10980
CGTACGATAC TTCTTAACTG GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTCTTGAG	11040
AAAACAAAAG GGAGAGTACC TGAGAGCGGA GAATGCCATC CTAaaaaaAGT TAAGAGAACT	11100
CCGATTGAAG GAGGAAAAAG AGAAATAAGA AAGACAGAAA TTGTTCAAGA ATTAATGACT	11160
GAGTTTTCGT TAGATCTTCT TCTAAAAGCC ATTAACTAG CTCGTTGGAC CTACTACTAT	11220
CACCTGAAAC AGCTAGATAA ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATCAATCC	11280
ATCTTTATCG AACACAAGGG AGATTATGCT TATCGCCGGG TTCATTTAGA ACTAAGAAAT	11340
CGTGCTTATC TGGTAAATCA TAAAAGAGTT CAAGGCTTGA TGAAAGTACT CAATTTACAA	11400
GCTAGAATGC GACAGAAACG AAAATATTCT TCTCATAAAG GAG	11443

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5338 base pairs

452

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

CCAATTACAT TATATTATCA AAATCGTCGA AACTGGCTCC ATGAATGAGG CAGCCAAGCA	60
ACTCTTTATC ACTCAGCCAA GTCTCTCCAA TGCAGTCCGA GATTTGAAA ATGAAATGGG	120
CATTGAGATC TTTATCCGCA ATCCCAAGGG AATCACCTTG ACCCGTGATG GCATGGAGTT	180
TCTCTCTTAT GCCCCTCAGG TTGTCGAGCA GACCCAGCTT CTGGAGGAAC GCTATAAAAA	240
TCCTGTCGCC CACCGCSAAC TCTTTAGCGT TTCGTCTCAA CACTATGCCT TTGTGGTCAA	300
TGCTTTTGTC TCTTTGCTCA AGAAAAGCGA TATGGAGAAA TACGAACTCT TCCTTCGTGA	360
AACTCGGACT TGGGAGATTA TCGACGACGT CAAGAACTTC CGCAGTGAGG TCGGGGTCCT	420
CTTCTTAAAC AGTTACAACC GTGATGTTTT AACCAAGATG CTGGATGACA ATCACCTGCT	480
AGCCCACCAT CTCTTCACAG CGCAACCGCA TATCTTTGTC AGCAAGACCA ACCCTCTGGC	540
AAAGAAAGAC AAGGTGAAAC TGTCTGATTT GGAGAATTTT CCTTACCTCA GCTATGACCA	600
AGGGACGCAC AACTCCTTCT ACTTTTCAGA AGAGATTCTT TCTCAAGAAC ACCACAAGAA	660
ATCCATTGTG GTCAGTGACC GTGCCACCCT CTTTAATCTC TTGATTGGTT TGGATGGTTA	720
TACCATTGCG ACAGGGATTT TGAACAGCAA CCTAAACGGA GACAATATCG TTTCTATCCC	780
ACTGGATATT GATGACCCGA TCGAGCTGGT CTATATCCAG CATGAGAAAA CCAGCCTATC	840
TAAGATGGGC GAACGCTTTA TAGACTATCT CCTAGAAGAA GTTCAGTTTG ATAGTTGAGA	900
AATGATAAGA ACCAATATGT AGGCTAGCAA CAACCTGCAC ATTGGTTCTT TTTACTTATA	960
ATTAAAAGTT TCCCCTGCCA ACTTATCAGC TAGCTTGGGA AAGAGAGTAT AAAACTTATG	1020
GGCTAGGTTT AACAAAATCG GGAGATTGAG TTCTCGTTTC TTTTTCCTA TAATCTTGAC	1080
AATCTTTTTA GCCACTGCAT CTGGTTCTAG CAGGAAGCGA TCAACCGATT TAAGATAAGT	1140
TCCATCTGGG TCGGCTTGGT CGAAAAATCC TGTACGGATT GGTCTTGGAT TGAATGTTGT	1200
CACATAGACT CCATAGGGCA TAAGTTCGAG TCGCAGAGCA TTTGAAAAAC CAATAGCCGC	1260
AAACTTGGTC GCTGAGTAAA GACTAGACTT GCCAGTAGCT ATTAGACCTG CCATGCTGAC	1320
GATGTTGATG ATATGCCCTT TGCTGCTTTC CTTTATACGA GCCGCAAGGT GACGAGACAG	1380
ATTCATCAGG GCAAAGGTAT TGACCTCAAA CATCTGGTGA ATATCTTTAT CAGCAATCTG	1440
GTCAAATCCC TCAAAAATCC CGTAACCAGC GTTGTTAATC AAGACATCAA TCTTGCCATA	1500
GCGGAGATAA AGATCAGTTA CCAGAGCTTC TAGGGCTGAA TCGTCCGTAA TATCAATTTT	1560

453

AATCAATTCT GCATGGGAAT AATTTCGGTA GAGTTGGGCT AATTTTCTCT TATTTCTACC	1620
AAGCAAGATG AGTTGGTCAT TGGGCAGGAG TTTGACCATT TCTTGAGCTA GACCACCGCT	1680
AGCTCCGGTA ATGAGAATAG TAGGCATACT TATCCTTTCT GTGACTGCTA GATTTCCACT	1740
TCTTCCAAGT CTTTGACCAC ATGGACATTT TCAAAAATTG TGGCAGCGTC TTTCTTGAGT	1800
TTGCTAATAT CTTTGTAGAG GAAACGGGCA CTGATATGGT TGAGTAGGAG GCGTTTGGCA	1860
CCTGCTTCTA CCGCTACTTG TGCAGCTTGC ATATTAGTTG AGTGACCATG GTTACGAGCA	1920
ATTTTTCAT CACCCTTGCC ATAAGTGGAC TCATGAAC TA GGACATCTGC ATTGACAGCC	1980
AGACGCACAC TGGCACCCTG TTTTCGAGTG TCTCCTAAAA TACTGATAAT CTTACCTGGA	2040
CGTGGCGCTG AGATATAGTC TGCTGCCTTG ATTTCAAGTTT CGTCTTCCAA AACAAAGATCC	2100
TGGCCGTTTT TGATTTTACC AAAAAGCGGG CCGAACGGAA CACCAGCAGC CTTGAGTTTT	2160
TCAGCATCCA GCCTCCCTTC TAGATCCTTT TGCATGACAC GATAGCCAAC ACAGAAAATA	2220
GTGTGGTCCA GCTCCTCTGC ATACACAGTG AATTTATCGG TTTCAAGAAT TTTACCCAGA	2280
GAATCTTGGT CAAACTCATG GAAATGAATG CCGTAGGGCA GACGAGAACC TGACACACGA	2340
AGGCTGGTTA AGACAAATGA CTTGATTCTT TGAGGTCCGT AGATTTCCAA ATCTGTCTGC	2400
TCTTCATTGG CCTGAAAGGC ACGGCTAGAA AGGAAACCTG GCAAACCAAA AATGTGGTCT	2460
CCATGCAGAT GGGTAATAAA GATTTTGCTG ACCTTACGTG GTCGAATTGT GGTTTCCAGA	2520
ATGCGATTTT GCGTACCTTC TCCACAGTCA AAGAGCCAAA CTTCTGTTAAT CTCATCCAAA	2580
AGTTTCAGGG CGAGACTTGA AACGTTGCGG GCTTTAGAGG GCTGACCAGC CCCCCTTCCT	2640
AAAAATTGAA TATCCATTCT ATACTTCTA ATTAATCAAT ATATAACATG GCTGTGCGGT	2700
TTTCCGATCG GAAATAGCGT TTGCCAGAAA AAGCAGCAGC TTCTTGCAAT AAATCCTCTT	2760
GGCTGTAGCC TTTGAGACGT TTTGACCAT CAGCCAATCT TTCCAAATCA GTCAAAGCTC	2820
TGAGACTTTC TAGGCTGATA ACTTCCTCGT CCTCGACAGG CTTCAATGTA ATCTTACCAG	2880
ACTCTTCAA GACTAATTGA TGGGGGAAAA TTTGCGCAAT TTCAAAGAGC AAGTCATCCG	2940
AGATTTTCTC CTCATTTTCA AAGAAAATCC GACCAAGGCC GTCACCTCTA TAACAAAAAC	3000
CAAAGGATTT ACCAGACAGA TTAAGCCGAA TAAAAGGCTT ATTTTCTAGG GTGAAACTTG	3060
GCTCAGTATT GTAAAGATTC AGTTCCTGAC TGAGTTCTGC AAAATAATCC GTCGCAGCCT	3120
GAGGACTCTT TTTCTGATAG AGTCTGCAA AGTAGGCATT AACAACACTT GCGGAGGTG	3180
TAATAAGTGT TAACTCCTCC TGATCTGTTT TACCAGCTAG AAGCTGATCC AGATAGACCT	3240
TGTCCAGACT TGTATAACCT CCATACCTTA GAGCCAAAGT TTTAATATCA GTCATAAAAT	3300

454

TCTTCTAACC	TCCATTTATT	TTTCTCGGAA	ATGTAGCCTG	TAAATCTTC	GCCGTCTTCC	3360
TGATAATCAC	GTTCTTCCAG	AATTGCAACA	CTCTCTAAAT	CATGAATCTT	GTAGGACTTT	3420
GAAAAAGGCA	CTCGCAGGGT	AAATGCTTCA	AAAATTTCTT	TAATCTTATC	TAGCAATAAT	3480
GCTTGCAAGT	TTTCACGACT	GTCTCAGAC	TTGGCAGAAA	TGAGGGTATA	TGGCGTTTGG	3540
GTAGGCGTGA	AATCCTCCAC	CAAATCCGCT	TTATTATAAA	GCGTCAAGTG	AGGAATATCT	3600
TCCATGTCCA	GGTCTTTCAT	GATGGAGAGA	ACCGTTTTTT	CATGCTCCTC	GTGGTAAGGA	3660
TTGCTAGCAT	CGATAACATG	AACCAGAAGG	TCCACATGCT	TGCTTCTTTC	CAAGGTTGAC	3720
TTGAAACTGG	ACACCAACTC	TGTCGGCAAA	TCTTGGATAA	AGCCAACGGT	ATCTGTCAAA	3780
GTTACTTGGA	GATTGCCTCC	CAGATGAATA	CTCTTGGTGT	TCCCATCCAG	AGTCGCAAAG	3840
AGCTCATCTG	CTTCATACTG	GGTCTTACTG	GTCAAGATGT	TCATGATAGT	TGATTTCCCA	3900
GCATTAGTAT	AACCAATCAA	ACCAATCTTA	AAAGTGCTAG	ACTCCAAACG	TTTTTCTCTG	3960
ACAGTCGCAC	GATTTTCTCT	AACCACCTTG	AGCTGGCGCT	CGATATCCGT	GATTTGATTG	4020
CGAACGCTAC	GACGGTTCAG	CTCCAGCTGG	CTTTCACCAG	GACCACGGGA	ACCAATTCCC	4080
CCTGCTGAC	GGCTGAGCAT	AATCCCCTGA	CCAACCAAGC	GAGGCAAAAG	GTATTTGAGT	4140
TGGGCTAGGT	GGACTTGGAG	CTTCCCTTCA	TGGCTTCGAG	CCCGCATGGC	AAAGATATCC	4200
AAAATCAACT	GCATACGGTC	AATGACCTTA	ACACCGAGAA	CTTCCTCTAG	ATTGACATTC	4260
TGCCTTGGGG	TCAGACGATT	GTTGACGATG	ACAGTAGTGA	TTTCTTCTGC	ATCCACCATA	4320
AGCGCAATCT	CTTCCAACTT	ACCAGAGCCG	ACGAAGGTCT	TGGAATCATA	TTTTTCACGT	4380
TTTTGTCTGT	AGCTATCTAC	AACGACTGCC	CCTGCCGTTT	TGCTAAACT	AGCCAATTCT	4440
TCCATGGAGA	GGTCAAAACT	GTCCATACCC	TGCAATTCCA	CACCAATCAG	CAGGACTCCG	4500
TCCTCTTTTT	TCTCCGTTTC	AATCATCTAA	AAACTCCTCT	ATCTGGCTTA	AAATGCGGTC	4560
TTGTACACCA	GATTCTCCAA	TCTGATAAAA	GGTGACCTGC	ATGCGATTAC	GGAACCAGGT	4620
CAGCTGACGC	TTGGCAAAAC	GACGAGTCGC	CTGTTTAAGA	CTCTCACTAG	CTTCCTCCAA	4680
GGTCTGCTCT	CCACGGAAAT	AAGGAAAGAG	TTCTTTATAG	CCAAATCCTT	TAGCAGCCTG	4740
TACATTAGGG	GAATGGTCAA	ACAGCCACTT	GGCCTCATCC	AAAAGCCCAG	CCTCAAACAT	4800
CAAATCCACT	CGGTGGTTGA	TACGCTCATA	AAGTTGACTA	CGTTCATCAT	CCAAGCAGAT	4860
AATCAGCGGT	TCATACAAGG	TCTCTTGATT	TTCCAAATCC	TGACCAAAAT	GGGCAATTTC	4920
TAAGGCACGC	ATAGCACGAC	GACGATTAAA	CTGGGGAATC	TCAAGGCCTG	CTTGATCCAC	4980
CAAATGGGCT	AATTCCTCAT	CTGAATATGG	CTCCAAACTA	GCTCGATAAG	CTAAAATCTC	5040
CTCATGAGGA	GTCTCCCCAC	CTAGGTGGTA	ACCTTCTAGC	AAGCTCTGGA	TATAAAGTCC	5100

455

AGTCCCACCG GCGATAATGG CTAGCTTGCC ACGGTTGTGA ATACCCTCAA TAGTCATCTT	5160
AGCTTCTGAA AAAAAATCAA AAGCCGAGTA AGACTCGGTT ATCTCTCTAA CATCGATTAA	5220
ATGATGAGGA ACAGCTGCCT GCTCTTCTGG ACTAGCCTTG GCCGTCCCAA TATCAAGTCC	5280
TCGATAGACT TGCTGGCTAT CTCCACTAAC CACTTCGCCA TTAAAACGCT TTGCGGGG	5338

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19446 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

CGGAAACCCA TCTAGTCTCC ATCGTTTGGG AGACCAAGCA ACACGAATCT TAGATGCTTC	60
TCGCCAACAG ATTGCAGATT TAATCGGTAA GAAAAGCGAT GAAATCTTCT TTACCTCGGG	120
TGGAACAGAA GGGGATAACT GGCTTATCAA GGGTGTGGCC TTTGAAAAAG CTCAGTTTGG	180
CAAGCACATC ATTGTTTTCAG CCATTGAACA TCCAGCAGTC AAAGAGTCAG CCCTCTGGTT	240
GAAAAGTCAA GGATTTGAAG TGGATTTTGC TCCAGTTGAT AAGAAAGGCT TGGTCGATGT	300
TGAGGCGTTA CAGGTTTGAT ACGGCATGAT ACAATCCTCG TTTCCATCAT GGCTGTGAAC	360
AATGAAATCG GCTCTATCCA ACCTATTGAG GCTATTTTCAG AATTCTTGGC AGACAAGCCG	420
ACTATTTCTT TCCACGTTGA TGCGGTTTTC GCGCTTGCCA AAATTCGGAC TGAAAAGTAT	480
CTGACAGAAC GGGTGGATTG CGCGACTTTC TCTAGTCACA AGTTCCACGG GGTTCGAGGT	540
GTTGGCTTTG TCTATATCAA ATCTGGCAAG AAGATTACAC CTCTTCTTAC AGGTGGTGGC	600
CAGGAGCGAG ATTATCGTTC GACAACTGAA AATGTGGCAG GGATTGCAGC GACAGCCAAG	660
GCCCTCCGTT TGTCTATGGA AAAGCTAGAT ATCTTTAGGA GCAAGACTGG GCAGATGAAG	720
GCAGTGATTC GCCAAGCTCT TCTGAACTAT CCGGATATTT TTGTCTTTTC AGATGAGGAA	780
AACTTTGCAC CTCATATTCT GACTTTTGGG ATCAAAGGTG TTCGAGGTGA AGTCATCGTT	840
CACGCCTTTG AAGACTATGA TATTTTCATC TCAACAACCT CAGCTTGTTT ATCTAAGGCA	900
GGAAAACCAG CCGGTACCTT GATTGCCATG GGAGTGGACA AAGATAAGGC CAAGTCAGCT	960
GTGCGTCTTA GCCTAGACTT GGAAAATGAT ATGAGTCAGG TCGAGCAGTT TTTGACCAAG	1020
TTAAAATTGA TTTACAATCA AACTAGAAAA GTAAGATAGG AGCATTCATG CAGTATTCAG	1080
AAATTATGAT TCGCTACGGA GAGTTGTCAA CCAAGGGTAA AAACCGTATG CGTTTCATCA	1140

456

ATAAACTTCG TAATAATATT TCGGACGTTT TGTCTATCTA TACCCAAGTT AAGGTAACAG	1200
CAGATCGCGA CCGTGCCAC GCTTACCTCA ATGGAGCTGA TTACACAGCA GTTGCAGAAT	1260
CTCTCAAACA AGTTTTTGA ATTCAAACT TTTCTCCTGT TTATAAGGTT GAAAAATCTG	1320
TAGAAGTTTT GAAGTCTTCT GTCCAAGAGA TTATGCGGA CATCTACAAG GAAGGTATGA	1380
CCTTTAAGAT TTCTAGCAAG CGTAGCGACC ACAACTTTGA ACTTGATAGT CGTGAACCTCA	1440
ACCAAACACT TGGAGGGGCT GTATTGGAAG CCATTCCAAA TGTGCAAGTT CAAATGAAAA	1500
GTCCTGACAT CAATCTTCAG GTGGAGATTC GTGAAGAAGC AGCCTATCTT TCTTATGAAA	1560
CCATTCGTGG GGCTGGTGGT TTGCCAGTTG GAACTTCAGG TAAAGGGATG CTCATGTTGT	1620
CAGGAGGGAT TGA CTCACCT GTAGCAGGTT ATCTTGCTCT TAAGCGTGG GTGATATCG	1680
AGGCAGTTCA CTTTGCTAGT CCACCATATA CTAGTCCTGG TGCCCTCAAG AAAGCGCAGG	1740
ACTTGACCCG TAAATTGACC AAGTTTGGCG GAAATATCCA GTTTATAGAG GTGCCTTTCA	1800
CAGAGATTCA AGAGGAAATC AAAGCCAAAG CGCCAGAAGC TTATTTGATG ACTCTAACTC	1860
GTCGCTTTAT GATGCGGATT ACTGACCGTA TTCGTGAGGT ACGAAATGGT TTGGTTATCA	1920
TCAATGGGGA AAGTCTAGGT CAAGTAGCCA GCCAAACCCT TGAAAGTATG AAGGCTATCA	1980
ATGCTGTTAC CAACACTCCC ATCATTCGTC CTGTGGTTAC CATGGACAAG TTGGAAATCA	2040
TTGACATCGC CCAGGAAATC GATACCTTTG ACATTTCAAT CCAACCGTTT GAAGACTGTT	2100
GTACCATTTT TGCACCAGAT CGTCCAAAA CAAATCCTAA AATTAAGAAT GCGGAGCAGT	2160
ACGAAGCGCG TATGGATGTT GAAGGCTTGG TTGAGCGAGC AGTGGCTGGA ATCATGATTA	2220
CTGAAATCAC ACCTCAAGCC GAAAAAGATG AAGTTGATGA CTTGATTGAC AATCTGCTCT	2280
AATTCAGAAA ATCCAAAAGA ATAGCGAAAA TCAGTAAAAA AAGTTAGTTT TTTCTCTAAA	2340
AACAGGTAAA AACTAACTT TTTTATTTT TATGATATAA TGATATAAAA TTTGAATAA	2400
AGAGAGTTT CTGACAATGA ATCAATCCTA CTTTATCTA AAAATGAAAG AACACAACT	2460
CAAGGTTCTT TATACAGGTA AGGAGCGCCG TGTACGTATT CTTCTTCCTA AAGATTATGA	2520
GAAAGATACA GACCGTCTCT ATCCTGTTGT ATACTTTCAT GACGGGCAAA ATGTTTTTAA	2580
TAGCAAAGAG TCTTTCATTG GACATTCATG GAAGATTATC CCAGCTATCA AACGAAATCC	2640
GGATATCAGT CGCATGATTG TCGTTGCTAT TGACAATGAT GGTATGGGGC GGATGAATGA	2700
GTATGCGGCT TGGAAGTTC AAGAATCTCC TATCCAGGG CAGCAGTTTG GTGGTAAGGG	2760
TGTGGAGTAT GCTCAGTTT TCATGGAGGT GGTCAAGCCT TTTATCGATG AGACCTATCG	2820
TACAAAAGCA GACTGCCAGC ATACGGCTAT GATTGGTTCC TCACTAGGAG GCAATATTAC	2880
CCAGTTTATC GGTTCGGAAT ACCAAGACCA AATTGGTTGC TTGGGCCTTT TTTTCATCTGC	2940

457

AAACTGGCTC CACCAAGAAG CCTTTAACCG CTATTTGAG TGCCAGAAAC TATCGCCTGA	3000
CCAGCGCATC TTCATCTATG TAGGAACAGA AGAAGCAGAT GATACAGACA AGACCTTGAT	3060
GGATGGCAAT ATCAAACAAG CCTATATCGA CTCGTCGCTT TGCTATTACC ATGATTTGAT	3120
AGCAGGGGGA GTACATCTGG ATAATCTTGT GCTAAAAGTT CAGTCTGGTG CCATCCATAG	3180
TGAAATCCCT TGGTCAGAAA ATCTACCAGA TTGTCTGAGA TTTTTCGAG AAAAATGGTA	3240
AGTTAAGAAA GGAAAAACG AAATGCATAT TGAACATCTT AGCCACTGGA GTGGTCATCT	3300
TAACCGTGAA ATGTACCTTA ACCGTTATGG ACATGGTGGG ATTCCAGTTG TGGTCTTTGC	3360
TTCATCAGGT GGTAGTCACA ACGAATACTA TGATTTTGGC ATGATTGATG CCTGTGCTTC	3420
CTTTATCGAG GAAGGCCTTG TCCAGTTCTT TACCCTATCT AGTTTGGATA GTGAGAGCTG	3480
GTTGGCTACT TGGAAAAATG CTCATGACCA AGCGGAAATG CACCGTGCCT ACGAACGTTA	3540
TGTGATTGAG GAGGCCATTC TTTTATCAAG CACAAGACAG GTTGGTTTGA TGGCATGATG	3600
ACGACAGGTT GCTCTATGGG AGCCTATCAT GCACTCAATT TCTTCCTCCA GCATCCAGAT	3660
GTCTTTACCA AAGTGATTGC TCTCAGTGGT GTTTACGACG CACGTTTCTT TGTGGGTGAT	3720
TACTACAACG ATGATGCTAT TTACCAAAAC TCGCCAGTAG ATTATATTG GAACCAAAAC	3780
GACGGCTGGT TTATTGACCG TTACCGTCAG GCAGAGATTG TGCTGTGTAC GGGGCTTGGA	3840
GCCTGGGAAC AAGATGGTTT GCCATCCTTT TACAAGCTCA AAGAAGCCTT TGACAAGAAA	3900
CAAATTCAG CCTGGTTTGC TGAATGGGGA CATGATGTCG CCCATGACTG GGAATGGTGG	3960
CGTAAACAAA TGCCCTATTT CCTCGGTAAT CTCTATTTAT AAAAGGAGTT ACCTATGAAT	4020
TACCTTGTTA TTTCTCCCTA CTATCCACAA AACTTTCAAC AGTTTACCAT CGAACTAGCT	4080
AATAAAGGCA TCACAGTCTT GGGAAATGGT CAAGAGTCTT ACGAGCAATT GGATGAGCCC	4140
TTGCGCAATA GCTTGACCGA GTATTTTCGT GTTGATAATC TTGAGAACAT AGATGAAGTC	4200
AAACGTGCAG TTGCTTTTCT CTTTTATAAA CATGGTCCAA TTGGCCGCAT CGAGTCTCAC	4260
AATGAATACT GGCTTGAGCT AGACGCAACA CTCAGAGAAC AATTCAATGT TTTTGGTGCC	4320
AAACCAGAGG ATCTCAAAAA GACGAAATAT AAGTCTGAAA TGAAGAACT TTTCAAAAAA	4380
GCAGGTGTTT CTGTGGTACC TGGAGCTGTT ATCAAGACGG AAGCAGATGT TGATCAAGCA	4440
GTGAAAGAAA TCGGTCTTCC AATGATTGCC AAACCTGATA ATGGAGTGGG AGCAGCCGCA	4500
ACCTTTAAAC TTGAGACAGA AGACGATATC AATCACTTCA AGCAAGAATG GGACCATTCA	4560
ACCCTTTATT TCTTTGAAAA ATTTGTCACT TCCAGCGAAA TCTGTACCTT TGACGGGCTC	4620
GTGGACAAGG ATGGAAAGAT TGTCTTCTCA ACAACCTTTG ACTACGCCCTA TACACCGCTT	4680

458

GACCTCATGA	TTTATAAGAT	GGACAATTCT	TATTATGTGC	TCAAGGATAT	GGATCCTAAA	4740
CTGCGCAAGT	ATGGGGAAGC	AATTGTCAAA	GAATTTGGTA	TGAAAGAACG	GTTTTTCCAT	4800
ATTGAGTTCT	TCCGTGAGGG	GGACGATTAT	ATTACCATCG	AGTACAATAA	CCGCCCTGCA	4860
GGTGGTTTTA	CCATTGATGT	TTATAACTTT	GCTCATTCCT	TGGACCTTTA	TCGTGGCTAT	4920
GCAGCTATTG	TCGCAGGAGA	GGAGTTCCCG	GCGTCAGACT	TTGAAACTCA	GTATTGTTTG	4980
GCTACTTCTC	GCCGTGCAAA	TGCTCACTAT	GTTTATTCAG	AAGAGGATTT	GCTTGCCAAA	5040
TATAGCCAGC	AGTTCAAGGT	TAAAAAAGTC	ATGCCAGCTG	CCTTCGCGGA	ACTTCAAGGA	5100
GATTACCTGT	ATATGCTGAC	CACTCCGAGT	CGACAAGAAA	TGGAGCAGAT	GATTGCAGAT	5160
TTCGGACAAC	GTCAAGAATA	AGAACTATCG	GATTAAGGAA	ATTAACTCCC	TTAATCCTTT	5220
TGTTTTGTCT	GATAAAAAAT	AAGAGCATCC	CAACAAGGTA	GCTATCATAA	AACTTGTTTCG	5280
ATAACTATTT	GAAGCAGGAT	TAGGTGGTCA	GAAATTAAT	TTAATATTT	CAATTGAGTC	5340
ATAGTATTGT	GTTTGGGTAT	CCTTAAATCA	GCTAAAAGGA	TCCATGACGA	CACCTATACG	5400
ATATAGTTTT	CAAGATACCA	AACAAGTCTA	TTAATATTCA	ATGAAAATCA	AAGAGCAAAC	5460
TAGGAAGCTA	GCCGCAGGTT	TCTCAAAACA	CTGTTTTGAG	GTGTGGGATA	GAAGTCACAG	5520
AGTCAGTATC	ATATACTACG	GCAAGGTGAA	GCTGACGTGG	TTTGAAGAGA	TTTTCGAAGA	5580
GTATAAAATA	TTCAGGTGAC	GCATAGATAT	AGTTAATTGA	AGCTTTGTTT	GAAATCTGAT	5640
AAAATAATGA	TATTACTAAG	TTTTAAAAAC	TAAAGAAAAG	GGAAGATATG	ATTACAGGCG	5700
AATTAaaaaa	TAAAAATGAT	CAGCTGTGGG	AAATTCTTTG	GACAGAAGGA	AACGCAAAATC	5760
CTTTAACAAA	TATTGAACAG	TTGACTTATC	TCTTATTTAT	GAAAGATTTG	GATAGTGTCTG	5820
AGCTTGGACG	TGAAAGTGAT	GCTGAATTTT	TAGGGATTCC	TTATGAGGGA	GTTTTTCCAA	5880
AAGATAAACC	TGAATACCGT	TGGTCAACTT	TTAAAAATAT	AGGAGATGCT	CAGGAAGTTT	5940
ATCGTTTAAT	GAATCAGGAG	ATTTTCCGT	TTATTAAAAA	TCTCAAGGGG	GATACAGATG	6000
ATACAGCCTT	TTCACGATAT	ATGCGAGAAG	CTATTTTTC	AATAAATAAA	CCTGCTACGC	6060
TTCAAAAGGC	AATTTCTATC	TTAGATGTTT	TTCCAAGTAG	GGGATTAGAT	GTAGATTTTG	6120
ATAATGACAA	ACAAAGTATT	ACTGATATCG	GAGATATCTA	TGAATATCTG	TTATCAAAAT	6180
TGTCGACCCG	AGGTAAAAAT	GGACAGTTCC	GTACACCTCG	TCACATCATC	GATATGATGG	6240
TTGAGTTGAT	GCAACCGACT	ATCAAAGATA	TCATCTCAGA	TCCCGCTATG	GGTTCGTCTG	6300
GCTTCTTAGT	ATCTGCTAGC	CGTTACTTAA	AGCGTAAGAA	AGATGAATGG	GAAACCAATA	6360
CAGATAATAT	CAATCATTTT	CATAATCAGA	TGTTTCATGG	AAATGATACG	GATACGACTA	6420
TGTTGAGACT	TGGGGCGATG	AACATGATGC	TACATGGAGT	AGAAAAATCCA	CAAATCAGTT	6480



459

ACCTTGACTC GCTGTCTCAA GATAATGAAG AAGCCGATAA ATATACTTTG GTTTTAGCAA	6540
ATCCTCCTTT TAAGGGCTCA CTTGACTACA ATTCAACCTC TAATGACCTT CTTGCAACCG	6600
TAAAAACCAA AAAACAGAA TTACTCTTTC TTTCTCTTTT CTTGCGAACT TTAACCAG	6660
GTGGACGAGC AGCAGTTATC GTACCTGATG GTGTCCTTTT TGGTTCGTCT AAAGCTCATA	6720
AAGGAATTCG TCAGGAAATT GTAGAGAATC ATAAGCTTGA TGCTGTAATC TCAATGCCTA	6780
GTGGTGTGTT CAAGCCTTAT GCTGGAGTTT CAACTGCCAT TCTCATCTTT AAAAAACTG	6840
GTAATGGTGG TACTGACAAA GTCTGGTTTT ACGATATGAA AGCGGATGGT TTAAGTTTGG	6900
ATGATAAGCG ACAACCGATT AGCGACAATG ATATTCCAGA TATTATCGAA CGCTTTCATC	6960
ATCTTGAAAA AGAAGCAGAA CGTCAGAGAA CGGATCAATC TTTCTTTGTT CCAGTTGCTG	7020
AGATAAAGGA AAATGATTAT GATTTGTCTA TCAATAAATA TAAAGAGATT GAGTATGAAA	7080
AAGTTGAGTA TGAACCAACA GAAGTCATAT TAAAGAAAAT CAATGATTTA GAAAAAGAAA	7140
TTCAAGCTGG CTTGGCTGAA TTGGAAAAAT TACTCAAGTA GGGAGGTGGC TGTATGAAAA	7200
AAGTGAAGTT GGGGAAGTC TTATCTCTAA AAAAAGGCAA GAAAGCCACT GTACTTGCTG	7260
AACAAACAAC TCTAAGCCAA CGTTATATTC AAATAGATGA TTTAAGAAAT AATAATAATT	7320
TAAATTCAC TGAAAGTTTA AATATGACTG AAGCACTCCC AGATGATATT CTGATAGCAT	7380
GGGATGGAGC TAATGCAGGA ACAGTTGCTT ATGGATTATC GGGAGCTGTT GGTAGTACAA	7440
TTACGGTCTT AAAAAAGAAT GAGCGATACA AAGAAAAAAT TATATCAGAT TACTTGGGAG	7500
TCTTTTGGGA AAGTAAATCG CAGTATTTAC GAGATCATTC AACAGGTGCA ACAATTCCTC	7560
ATTTAAACAA GAATATATTA CTTGATTTAC AATTAGAATT GCTAGGTATC GAAGAACAAG	7620
AGAACATTAT CTGTATCTTT AATACGATTA AAAGGCTTAT TACTAAAAGA AAATTCAGT	7680
TAGATGAACT AAACCTGCTC GTCAAAATCCC GATTTAACGA GATGTTTGGG GAAAATAAAA	7740
TATTTGAAAG CATTGATAAC TTATTTGATA TTATAGATCG TGATAGGGGC AAAAATTATC	7800
CTAAATCAGA TGAGTTGTTT AGTGAGGAGT ACTGTTTATT TTTAAATACA AAGAATGTTA	7860
CTAAAAACGG ATTTTCATTC GATACAAAGC AATTATCAC TAAAACAAAG GATAAATTAC	7920
TTCGAAAAGG CAACTTGAG CGTTATGATA TAGTCTTGAC AACAAGAGGT ACTGTTGGAA	7980
ATGTAGCGTA CTACGATGAA TTAATAAAAT ATAAACATT ACGTATAAAT TCAGGTATGG	8040
TAATATTACG TCCCAAGACA CCAAATCTAA ATCAGAAAT TATTATCCAT GTTTTAAGGA	8100
ATAATAATTA TAGTCGAGTG ATATCAGGAA GTGCTCAGCC TCAGTTACCA ATTACAAAAT	8160
TAAAAAAAT ACTTCTCCCC CTCCCCCAC TAGCCCTCCA AAATGAGTTC GCAGACTTTG	8220

460

TAGTCCAGGT CGACAAATCA CAATTGCTT GTGAGATAGC TATAAAAGTG TGGAGAAATA	8280
GCTTGAAATT TAGTATAATA TAGCTAAACT ATTTGTTTAA AGTGAGAAAA AAATGGGAAA	8340
TTTAGCTTT CTTTAAAAA ATGACGAATA TGAATCTTT TCAAAACCTT GCATTGAAGC	8400
TGAGAATATG ATTGCTACAT CAACTGTGGC TACTGCCTTT ATGGCGCGTC GTGCTTTAGA	8460
GCAGGCTGTC CATTGGATAT ATAGTCACGA TTCATATTTA GAAGCTCCCT ATCGTGTCTAC	8520
TCTATCTTCT TTAGTATGGG ATGATGATTT TAGGGATATC GTAGATTCTG AACTCCACAA	8580
GCAGATAGTT CTGTTGATTC GGTGGGAAA CCATGCTGCT CATGGTGGTG AAATTAAGGA	8640
ACGAGAAGCG ATTTTAGCTT TGCATCATTT GTATCAGTTT GTTAATTTTA TCGATTATTG	8700
TTACAGCAAT GAGTTTGTGG AGCGTTATTT TGATGAGAAG TGCTTACCAC TTTCAGCAA	8760
CATCAAATAC CGAGAAACTC CACAATCTAT GATAAAGTTA CAAGACAGTT TACCAGAACT	8820
GCCTGATTTT CATGAACAGA TGGCTGCTCA GTCGGTAGAA GTTCAAGAGA CTTATACTGA	8880
AAAACGTGAG ACTGCAGCGC AACGGCAAGA TGTGCCTTTC CATATTGATC AATTATCTGA	8940
GGCAGAGACA AGAAAGCTCT TTATTGATAT CGATCTCCGT TTAGCAGGAT GGATATTTGA	9000
AGAAAACTGT CGTGTGAGA TAGCCGTTGA TGGTCTCAAG CACGGTTCAG GAATTGGTTA	9060
CTGTGACTAT GTACTTTATG GTAAAAATGG GAAAAATTTA GCGATTGTGG AGGCTAAAAA	9120
AGCCTCTGTC AATCCAGAAG TAGGGGAAGT ACAGGTCAAA GAATATGCTG AAGCTTTGGA	9180
GAAACATATC GGCTATCAGC CAATTGCTT TATTACAAAT GGTTGAAGC ACTATATACT	9240
TGATGGTCCG AACCGCCGCC AGATTGCAGG CTTTACTCT CAAGAAGAAT TGCAATTAGT	9300
GATGGATAGA CGTCATCTTC AAAAACCGCT TGAGGATATT TCTAGTAAAA TTAGGGACGA	9360
TATTTCCGGG CGTCACTACC AAAACATGC CATTCGAAGC GTTTGTGAAG CTTTCTCTGA	9420
TCATCGTAGA CAGGCACTTT TGGTTATGGC AACTGGGGCG GGGAAAACTC GTACAGCAGT	9480
TTCTCTAGTT GATATCTTAT CACGTCATAA CTGGGTAAAA AACGTTCTCT TCTTAGCCGA	9540
TAGAACTTCC TTGGTTAAGC AAGCATATGA TTCGTTTGA AAATTACTCC CAGATCTTTC	9600
CGTTTGTAAC TTCTTAGAAG ATAAAGAAGG AGCTCAATCA AGTCGCATGG TCTTTTCAAC	9660
TTATCCGACC ATGATTGGAG CGATTAGTGG TCAAGAAGAA GTAAATCAAC GCCCTTTCAC	9720
TGTTGGGCAT TTTGACCTTA TCATAATTGA CGAATCTCAC CGTTCTATTT ATCAGAAATA	9780
CAAGTCCATT TTTGATTATT TTGATGCAAG AATTGTAGGC TTAACAGCTA CTCCGCGTCA	9840
AGATTTAGAT AAAACACCT ATGGATTCTT TAATTGGAG AATGGGGTTC CAACATATGC	9900
ATATGATTTG GAAGAGGCTG TTAAAGACGG ATATTTAGTA GCCTATCATT CTATCGAAAC	9960
CAAAGTAAA CTACCTACGG ATGGTCTACA TTATGATGAT TTGTCCGAAG AAGAAAAGGA	10020

461

ACATTTTGAT AGCAAATTTG AAGACAATAG CTGTGAAAAA GATATTGATG GGAGTGATT	10080
TAATTCCTTT ATTTTCAATA AAAGTACAGT AGAAATTGTT TTAAATGAAC TCATGACAAG	10140
AGGAATTCAG ACAGCCTCGG GTGATGAAAT TGGTAAAACT ATTATTTTGG CTA AAAATCA	10200
TGATCATGCG GAATATATCA GAGGTATTTT TAACAACCGC TATCCTGAAA AAGGGAGCGA	10260
CTATGCTCAG GTGATTGATT ATAGTATTAA GCATTATCAG ACCTTGATTG ATGATTTTAA	10320
AATTAAGGAG AAGTATCCTC AAATTGCGAT TTCTGTGGAT ATGTTAGATA CAGGTATTGA	10380
TGTACCAGAG GTTGTTAATT TAGTCTTCTT CAAGAAAGTA CGCTCTAAAA CTAAGTTTTG	10440
GCAGATGATT GGTGAGGAA CCCGTCTATG TAAAGATTTA TTTGGACCTG AGCAGGATAA	10500
GGAAAACTTC TTGGTATTTG ATTATGGGGA CAATTTTGAT TATTTTCGTG CAGATCCAAG	10560
AGATGGAGAG GGTGCTCACA TTGTTTCGCT GACTCAGCCT TTATTTAATA TCAAAGTGG	10620
CTTGATTGGA GAACCTCAGG GACTCCAATA CCAAGAAGAT CAGTTTGCGA GAGCATACCG	10680
TCAGCAGCTT GTCTCGGAAC TTCAAGGTCG TATAGAGAGC TTAAATCAGT TGGACTTCAG	10740
GGTTCGTATG GTTTTAGATA CAGTTTATAG CTATAGGAAA TTGGAAAGTT GCCAGAATCT	10800
AACTGCTGTT ACAAGTGAAA CCATTCAAAA AAATCTCTCT CCGCTTTTAT TTGATGAAGA	10860
TAAAGAAGAT GAGATGGCGA GGAGATTGTA TTTGTGGTTG CTTTCATATC AGTTGGGGCA	10920
ACTGACAGCT AAATCTTCCA CTGTTTCATAT TTCCCAAGTG ATGAAGACGG CTAGAGCTCT	10980
TTCTGCTATT GGCAATATCC CGCAGGTTTT TGAGCAGGCT GAAATTATCA GGAAAGTACA	11040
GGAGCCTGAA TTTTGGAAG AAGTTAACTT GTCTGATTTG GAAAAATTC GTCTTGCTAT	11100
TCGAGATTTA TTACAGTTTT TGGATAAAAC AGACCGTAAA CCCTACTATG TTAACTTTGA	11160
AGATCGTATA CTCTCCACTG TTCACGAGAC CACAGCATT TGGCAGGTCA ACGATCTTCG	11220
GTCTTACAAT GAAAAAGTTG AGCATTATTT GAAAACTCAT CTGGATGAGG AGTCCATTT	11280
TAAGCTATAC CATAATAAAA AGTTGACATC TGATGATATG CTTGCACTTG AAAAATTGCT	11340
TTGGGAAAAA TTAGGTAGTA AAGCAGACTA CCAAAGTCAT TATGAAAATA AGGCAATTCC	11400
GAGATTGGTT CGTGAGATTA TTGGCTTAGA TAGAGAGTCT GCCAATCGTA TTTTCTCTAA	11460
ATTTTGTGCG GATGAGAATC TTAATGCCAG GCAGATTTC TTTGTAAAAT TGATTGTAGA	11520
CTACATTGTA GAAAATGGTT TTTTAGAGAC GAAAGTGTTA ACGCAAGAGC CGTTTAAATC	11580
TTATGGTTCT GTTCAACTAC TCTTCCAACA CCAACTACCA GTACTTCGTA ATATTGTTC	11640
AATCATTGAA CTTATCAATA ATCGAGCTGG AGAAGCGGCT TAAATCTTAA AGTGATTGCC	11700
ATGCTGAGAC TCATTTAAAA TTA AAAAGAG TAGAAATTTA TGCTATATAT GAGAAGTTTT	11760

462

ATTAGGAAGA ATGTCATCGT TTTCTAGAA TACAGTATCA GTTCTTAAGT GGTGATAAA	11820
TTTCAAAGTA GATACTTGTA CCACGATGTT TGTGATCGA GTTATTAACA AAAGAGCTAC	11880
TTTGATTTTA AAGAAATAGA AAACAAAAG CCGAGCAAGA ATTCAATTGC AGGAGAAAAT	11940
GAAATAATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAA CTAGCTGCAG GCTGCTCAAA	12000
ACACTGTTTT GAGGTTGCAG ATGGAAGCTG ACGCGGATTG AAGAGATTTT CGAAGAGTAT	12060
AAATCTTCCT AGGATAAAGC AAAACGCATA GTATCAAGGG TTTCAACAC TTGATACTAT	12120
GCGTTTTCTG ATGTTAAAGA CTTTCTACCA GGTTTTTTAA AAGCATAATT GTTAGTTGTA	12180
GTCAATTTATT ATTCTTCAAA GAAAAATGGT GGGGCGAATT TTTTCAGTTC TTCAAAGCAC	12240
TTTTGAGCAG TATCTGCATC TTCACAGATG ATAAGACAGA CATCAATTACC ACAAAGGGTA	12300
GCGATAGCGT CAGGGAAGCT CAAAGTATCA ATGATAGAAC CAAAGGATTG AGCCAGTCCA	12360
GGAAGGGTTT TTAGTAGGAC TTGGTGTGA ACTGGGCGCA TCCAGACAAG GGCCTCTTCC	12420
ATGTAGAGTT CGAGACGTTT TTCCCATTTT GAGATGGAAC CATTGTTAAG AACATAATAA	12480
GCGCTATCTT CTTGCGGAC TTTTGATAGG TTCATATTTT TGATGTCGCG TGAGAGGGTT	12540
GCCTGGGTTA CTTGAATGTC GTTCTCAGCA AGAAGGGCTT GCAACTCAGC CTGTGTATGA	12600
ATCTTGTTTT TTGTGATAAG AGCGCGTATA AGTTGGTGGC GGTGTTCTGA TTTATTCATA	12660
ATAATGTAAC TCCTTTTAGC AAGGTAAGGT AAGCATGGAC TGAGCGAGGT CGACAGTCAA	12720
GTGGTAGTCT GTATTGTCAC GGATGGTGAT TTCAAAGTCA GTAGTATAGA GGAATAAACC	12780
GAGAGTGTCT CCTTCTTTTA GCTTGTAAT AGTTGGCTGC AGTTCAAAT GAACGTCCAT	12840
CCATTCATCT GCAGTAATAT CCTCTACTAA CAGTAAATCA TTTCTATTTT GTAAATTAAG	12900
GTAACCTTTT GTCACGACTC GTTGTGCCTC TGGTCTAAAT GGCAATTCAC AGAGATTTTC	12960
CAACATGTGA TAGCGACCGT TGTCAATGGT TCTAGCACTT AAAATAGCTG GATAAGGTTG	13020
TAGGTATTTT TTTTGCCCAA ATTCTAGCAG TTGGGCAGAT AAGAGCCCCT TGTTTGTACT	13080
GGATTTGATA CGAAGATTGA GCTGAGCGCG ACCGTTTAGG TGGAGATCTT TAGTCACAGG	13140
AAGGTTAATA GTAATCTGAT TGGCTTTCCC TTGATAGAGC TCTGTATTGA AGGTTTGGTA	13200
TGTCTTACCA TAGCGCTCAA AATCCTTATC TGGGTACTGG TTTTGAATAG CTTGCTCTTC	13260
TTGACCAAGT GAGAAGGTTT CACAGTTTTC TTGCCCACCG AAGTTATCAA GTGATAACCA	13320
AGTCTGTGGA GCTGTATTGT CCTGCCAGAT AACAGTAGGA AGTTGAAAGT CTGTTTCCCTG	13380
TCCTAGTAAT TTCTTGTCAT ATAAGGCATT TATGCACTCA GGAAGTCAA TTGATTGCCA	13440
ATTGTTTCATG TAAACATGGG CACCATTATG GAAAAAGAGA TGCTTGCTTA TATGAGTAGG	13500
AAGAGCATGG AACATCTGGT AAACATGAAG TGGTTTGACA TTCCAATCCT GAGAACCATG	13560

463

AGTAAAGACA ACCTCTGCCT TTACTTTATG GGCATTGAGC AGATAGATGC GGTGATGCCA	13620
AAACTGATTG TAGTCCCCAG TTTTTCGGTC TAGCTGAGCT TTCACCTTTT CTAAGTCAGC	13680
TTGGTGAGCT TCATTGCCAC GGATATAGTC GCCAGCTAAG AGATTACGAG AATAGGTTAA	13740
CTCAGCAAGG GAGTCAAAGT CCTCACCTGG ATAACCACCT GGGCTAGTCA CCAGACCGTT	13800
TTCACGGTAG TAGTTGTACC ATGATGAAAT TCCTGCCTCG GCAATGATAA CTTCTAAACC	13860
ATCGACTCCT GTAGTCGCAA GACCATTGGA CATGGTACCT AGATAGGAAA GTCCTGTTGT	13920
AGCAACTTTT CCGTTTGACC AATCAGCCTT GACTTGACGC TGGCGCGTGT GATCAGTAAA	13980
GGCACGGCAA CGACCGTTAA GCCAATCGAT GACATTTTTA TAAGCCTCGA TTTGCTGGTA	14040
GTCTCCATTA GTCATGAAAC CTGTCGAGTC TTTGGTACCA ACACCTGAGA CATAGAGATT	14100
GGCAAAGCCT CTCGGAAGGA AGTAGTCGTT TAGTGTATAG CTAGAGTTGA TGTGAGTTAG	14160
CTTTTCCTCA GCCTCTGCTA TAAGCTCAGC TTTACCTTGG GGTGGACGA GATTTAGTTG	14220
AGGTTTCTCT AGCTCAATCT TGTGAGGAAG CTTAACCTCA AGCTCGCCCT CCATCTTGTA	14280
GAGAGCCTTG TCACTAGCCT TGTCATTGGT TCCCTGATGA TAAGGGCTGG CTGTCATGAT	14340
GGCAGGGATT TTTCCATCAA AACGAGGGCG AATAATGCTA ACCTTTACTA GGTCTGATAG	14400
CCCTTTTGG TCAGTATCGA CACGAGACTC AACGTAAACG ACTTCACGAA TGACATCCTG	14460
GTTAGAAAAA GTAGCCAAAC TCTTGCCGTT AAAGTAGTGG TAGTCATTAT CCTCCGGAAT	14520
AAGACCATCA CTAACAAGTT GGTGATAAG AGTATTTCTT TTTTGGTGC GAGTATTGAG	14580
TAAGTGATAG AGATTTTCAA TCAAGTCACC ATATATAATG GGAAATCCAG TTTCTTTACG	14640
AAAAACGTC CTATCTTCGA AGTCAACCAA ATAAGAAAAG CCTAAAAGTT GAAAAGCAAC	14700
AGTATAAAAA ATATCTCTCG TCAGTTCATC TCTGATTGA AAAAATGTC GCAGGTCGT	14760
TTTTTTATCA GCTGCTAGGA TAGAAAGTGG GTAGTTGGTG TCTTGATAAG TGAAAAAGAA	14820
ACGACGTAAA AAGGTTTCAA GTGAGTCTTT GTGATTGGCT GTATTTTGTA AATCAAAGCC	14880
ACATTTTTTT AGTTCAGATA AGACATTTTC TTTTGGAAAA TTGATATAAC TATATTGATT	14940
AAAACGCATA GAACCTCCAT ATAGAATGAC AGTTAAGGTT ATTATATCAA AAAAAAGCA	15000
GAAAGGGAAT TGTTAACTTC AAAAGGAAAT AATCCAATAA AAATGAATAA AGTACTAAAT	15060
TCAATATAGA GAACAGAGTA ACAATAAGAA TAAATAGATA GGTATAAAA GTTCTAGGAG	15120
ATTTATATTA TATGCTTTCT ATTTTATAT ACAATATAGT ATAAATATAA AAATGATGAC	15180
AAAAATACAA ATGAATAGAA AATAAATTAG TAAGCTGATG AAATTTTCT CAAGAGAAGC	15240
CATTTATAGG TGAAAATGGT ATAATATAGT GAGAAGGATA GAGGAGAAGT GTAAATTGAT	15300

464

CGCACAACTA GATACAAAAA CAGTCTATAG TTTTATGGAA AGCGTCATTT CGATCGAAAA	15360
GTATGTGAGA GCAGCTAAAG AATACGGCTA CACTCATTTG GCTATGATGG ATATTGACAA	15420
TCTTTATGGC GCTTTCGACT TTCTAGAGAT TACAAAAAA TACGGCATTC ATCCTTTGCT	15480
AGGGCTTGAA ATGACAGTGT TTGTAGATGA TCAGGGAGTG AATTTGCGCT TTTTAGCTCT	15540
ATCTAGTGTG GGCTATCAGC AGTTGATGAA GCTTTCGACA GCCAAGATGC AGGGGGAGAA	15600
AACTTGGTCA GTCCTGTCCC AGTACCTGGA GGATATCGCG GTCATTGTGC CTTATTTTGA	15660
TAGAGTTGAG TCGTTAGAAC TAGGCTGTGA TTACTATATA GGGGTTTATC CAGAAACACT	15720
AGCAAGCGAA TTTTCATCATC CTATCTTACC TCTTTATCGG GTCAACGCTT TTGAAAGCAG	15780
GGATAGAGAA GTTCTTCAAG TTTTAACAGC GATTAAAGAA AATCTACCGC TCAGAGAAGT	15840
TCCCTTGCGT TCGAGACAAG ATGTCTTTAT ATCAGCAAGT TCTTTAGAGA AACTATTCCA	15900
AGAGCGTTTT CCGCAAGCTT TGGACAATTT AGAAAAGCTT ATTTTCAGGCA TTTCTTACGA	15960
CTTGGA TACT AGTCTGAAAC TGCCTCGTTT TAATCCAGCT AGACCAGCAG TAGAGGAGTT	16020
GAGAGAGCGT GCTGAACTGG GGCTTGTTCA GAAGGGGTTG ACTAGTAAAG AATATCAAGA	16080
TAGACTAGAC CAAGAATTGT CTGTTATTCA TGATATGGGC TTTGATGATT ATTTCTTGGT	16140
TGTTTGGGAT TTGTTGCCTT TTGGACAATC GAATGGCTAT TATATGGGAA TGGGAAGGGG	16200
TTCTGCACTA GGCAGTTTGG TTTCTTATGC CTTAGACATC ACGGGGATTG ACCCAGTAGA	16260
GAAAAATCTG ATTTTGAAC GCTTTCTTAA TCGTGAACGC TATACCATGC CTGATATTGA	16320
TATTGATATC CCAGATATTT ATCGTCCAGA TTTTATCAGA TATGTTGGTA ATAAATATGG	16380
TAGTAAACAT GCGGCACAAA TCGTTACTTT TTCAACCTTT GGAGCCAAGC AAGCTCTTCG	16440
AGATGTCTTG AAACGCTTTG GTGTGCCAGA GTATGAATTA TCTGCAATTA CTAAGAAAAAT	16500
CAGTTTTCGT GACAACTCTA AGTCGGCCTA TGAGGGAAAT CTCCAGTTTC GTCAGCAAAT	16560
CAATAGTAAG TTAGAATACC AAAAAGCTTT TGAGATTGCT TGCAAGATAG AGGGCTATCC	16620
AAGGCAAACC TCTGTCCATG CGGCTGGTGT TGTAATTAGT GACCAAGATT TAACCAACTA	16680
CATTCTCTA AAGTATGGTG ATGAAATCC ACTGACTCAG TATGATGCTC ATGGAGTTGA	16740
GGCTAGCGGA CTTTGAAGA TGGACTTTCT GGGACTACGA AATTTGACCT TTGTCCAGAA	16800
GATGCAAGAG TTGCTTGCTG AAACAGAAGG TATTCATCTG AAAATTGAAG AAATCGATTT	16860
ACAAGACAAA GAAACGTTAG CTTTATTTGC CTCTGGTAAT ACAAAGGTA TCTTTCAATT	16920
TGAGCAACCA GGTGCCATTC GTCTGCTTAA GCGTGTGCAA CCAGTCTGTT TTGAAGATGT	16980
CGTCGCGACT ACTTCTCTAA ATCGACCGGG TGCTAGTGAC TATATCAATA ATTTTGTTGGC	17040
AAGAAAGCAT GGGCAGGAAG AAGTGACTGT TCTGGATCCA GTACTGGAGG ATATTTTGGC	17100

465

TCCAACCTAC GGCATAATGC TCTATCAGGA GCAGGTATG CAGGTTGCCC AGCGACTTGC	17160
CGGATTTAGT CTTGGGAAAG CCGATATTTT GCGTCGGGCT ATGGGGAAAA AGGATGCCTC	17220
TGCCATGCAT GAGATGAGGG CTCCTTTAT TCAAGGTTCA TTAGAAGCTG GTCATACTGT	17280
GGAAAAAGCA GAGCAGGTCT TTGATGTTAT GGAGAAGTTT GCAGGTATG GTTTTAACAG	17340
GTCACACGCC TATGCCTACT CAGCCTTGGC CTTCAGTTG GCTTATTTCA AAACGCATTA	17400
TCCAGCCATT TTTATCAGG TCATGTTAAA TTCTTCCAAC AGTGATTACT TAATAGATGC	17460
ACTTGAAGCA GGTTTGAAG TAGCCTCTCT ATCCATCAAC ACCATTCCCT ATCACGATAA	17520
AATTGCCAAC AAGGCCATCT ATCTAGGTTT GAAATCCATT AAAGGAGTCA GTAATGATTT	17580
AGCTCTCTGG ATTATTGAAA ATAGACCTTA TTCTAACATT GAAGATTTTA TAGCTAAATT	17640
ACCTGAGAAT TATCTGAAAC TTCTCTGCT AGAACCTTTG GTAAAAGTTG GTCTTTTCGA	17700
TTCAATTTGAA AAAAATCGTC AAAAAGTATT TAATAACTTA GCTAATCTAT TTGAATTTGT	17760
GAAAGAGTTG GGAAGTTTGT TTGGAGATGC TATTTATAGT TGGCAGGAAT CGGAAGATTG	17820
GACGGAACAA GAAAAATTTT ATATGGAACA AGAGCTTTTA GGGATAGGTG TCAGCAAACA	17880
TCCACTACAA GCTATTGCAA GTAAGGCTAT TTACCCGATT ACCCCAATCG GAAATTTGTC	17940
AGAAAAATAGC TATGCTATTA TCTTGGTTGA AGTTCAGAAA ATAAAAGTGA TTCGTACCAA	18000
AAAGGGTGAA AATATGGCCT TCTTACAGGC AGATGATAGT AAGAAAAAAT TGGATGTCAC	18060
TCTCTTTTCA GACTTATATC GTCAGGTTGG ACAGGAAATA AAAGAGGGAG CCTTCTACTA	18120
TGTAAGGGA AAAATACAAT CACGTGATGG CCGTCTGCAA ATGATTGCAC AAGAAATAAG	18180
AGAAGCAGTT GCTGAACGCT TTTGGATACA GGTGAAAAAT CATGAATCGG ATCAAGAAAT	18240
TTACCGCATT TTAGAACAAT TTAAAGGCC C AATCCAGTC ATCATCCGGT ATGAAGAGGA	18300
ACAGAAAACC ATCGTTTCTC CCCATCATTT TGTAGCTAAA TCCAATGAAT TAGAGGAGAA	18360
ATTGAATGAA ATCGTTATGA AAACGATTTA TCGCTAAAAA TACGGAAAAT AGAAGAATTT	18420
TCAACGTAAA TGTGGTATAA TCAGTAAGAA TGTAAAAAGA AAAAGGAGCA TAACCAATAT	18480
GAAACGTATT GCTGTTTGA CTAGTGGTGG AGACGCCCT GGTATGAACG CTGCCATCCG	18540
TGCAGTTGTT CGTCAAGCAA TTTCAGAAGG AATGGAAGTT TTTGGTATCT ATGACGGATA	18600
TGCTGGTATG GTTGCCGGTG AAATTCATCC CCTAGATGCA GCTTCAGTAG GGGACATCAT	18660
TTCTCGTGGT GGTACTTTCC TTCCTCAGC TCGTTACCCA GAGTTCGCTC AACTTGAAGG	18720
GCAACTTAAA GGGATTGAGC AATTGAAAAA ACACGGAATT GAAGGTGTAG TTGTTATCCG	18780
TGGTGACGGA TCTTACCACG GCGCTATGCG TTTGACTGAA CATGGCTTCC CAGCTATTGG	18840

466

TCTTCCAGGT ACAATCGATA ACGATATCGT TGGTACTGAC TTTAC, TCG GTTTTGACAC	18900
AGCGGTTACT ACTGCCATGG ACGCTATCGA TAAGATTTCGT GATACATCAT CAAGTCACCG	18950
TCGTACTTTT GTAATCGAAG TTATGGGACG TAACGCTGGT GATATCGCTC TTTGGGCTGG	19020
TATTGCAACT GGTGCTGATG AAATCATCAT CCCTGAAGCA GGCTTCAAGA TGGGAAGATAT	19080
CGTAGCAAGC ATCAAAGCTG GTTATGAATG TGGTAAAAA CACAATATTA TCGTCTTAGC	19140
TGAAGGTGTG ATGTCAGCGG CTGAATTTGG TCAAAAACTT AAAGAAGCTG GAGATACAAG	19200
CGACCTTCGT GTAACAGAAC TTGGACATAT TCAACGTGGT GGTCTCCTCA CTGCGCGTGA	19260
CCGTGTTTTG GCGTCACGTA TGGGTGCACA TGCTGTAAAA CTTCTTAAAG AAGGTATCGG	19320
TGGTGTGCG GTTGCTATTC GTAACGAAAA AATGGTTGAA AATCCAATTC TTGGTACTGC	19380
AGAAGAAGGG GCATTGTTTA GCCTTACTGC AGAAGGTAAG ATTGTGGTTA ACAACCCAGC	19440
TACAAA	19446

## (2) INFORMATION FOR SEQ ID NO: 52:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16593 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

TCGTAAATAT GCTCTGTTTT TGGATTTTGT TTCTTAATCT GTTGGCAAG TGCCTTCATC	60
ATAGAAATAG GACCACACAT ATAGACGGTT GCATGTTCCG GCACCTCTTT TTGTTCAAAA	120
TTAAGATAGC CGTCTTTCGT ACTGTCGATT AGATGGAGTT CAAAATTAGG ATTTTCTGA	180
GCATAGTTAC GGAGTAAATC TAGGTAGACT GCATTTTCAT CTCCACGGAA GCTATAGTAG	240
AAGTGAACCT GTTTATCTAA AATAGGATGT TCACGGATGT AAGAGATGAA GGGGGTGATC	300
CCAATACCTC CAGCAATCCA AACCTGATTT TCTCGTCCTT CTTCTATGAT CATGTGTCGG	360
TAAGCTCTGT CTAGGGTTAC TTTGCTGCCG GCTTGAAGAT TATCATAGAT ATTCTTGGTA	420
TGGTCGCCCTG AAGTTTAAAC AGTAAAGTAA AGAGTTTGAC CATGACCTCC TGAGATAGAA	480
AAGGGATGCG GAGCACTTTC AAAGCCTTCT TGGAAAATCT TTAGAAAGGC AAATTGTCCT	540
GATTGATAGT TGAAAGGTCT GCTAAGATGG ATTTGAATTT CTCTAGTATC GTGATTTAAG	600
CGTTTGAGAT GGGTAATTTT CCCTAGATAG GGGAAGGAAA TCTTTTGATA TAGAAAAATG	660
ATATAAAAC CAGCTAGTAA GCCTAAAAGG GCATAGCTAC CAACAAGAAA ACTTAGAAGA	720
TTAAATGTAA GGAGACGATT GCCCATTATC ATGTAGATGT GAAAGAGTCC TAAAATATAG	780



467

GCTAGGTAAA CCAGGCGGTG AATCCATCGC CAAGCTTCGT ATTGGATGTA TTTGCCTAAA	840
TAGGCGACAA GGATGATGCT GGCAAAGATA TAGATGGCAA GATTGCCAAA CTGAGCAGCT	900
AAGCGAGAGC CCCACAAACC GCCCATACTA AAGTTATGAA AGATTAGTAG GATGATTGAG	960
AGAAAGGCTG TGAATTTGTG GACGGTGTAG ACCTTCTCCA AACTGTGAAA CCAGCTTTCT	1020
AGTAGTGGGA GACGAGTGGC TAGGATAAAA GTCAGAGATA GGCTTGTTAA AGCTAGTCCT	1080
GGAATCATGA ATTGGGGAGA AGTGTTCATC CAAGTCAAAA GAGTCAAGAT AAAACTAGCT	1140
ATGATAAAGA GTAGTCCTTT GACTGATTTT ATAGAAAATT CCATTTCATT TAGATTTCSA	1200
TTTGTGTAA ATAAATTTGT TACATTTTAT CATAGAAAAT GTATGGTGTG AAATTGAGGT	1260
CTATAAATAT CTACTCTCAT CAAAAAACTC TCCAATTGAA CTGGAGAGTG GCTGTTTATA	1320
CTCAATGAAA ATCAAAGAGC AAAGTAGGAA GCTAGCCGCA AGTTGCTCAA AACACTGTTT	1380
TGAGGTGCA GATAGAGCTG ACGTGGTTTG AAGAGATTTT CGAAGAGTGT TATTCTGCAG	1440
CTTGTGCCA ACGTTTGGCT AGCATATGAG ACAGGCTAGA AATTGCTAGG TTAAGCTGA	1500
AGTAGATGAG GGCAATCAGG ATGTAAAGAC TGAAGACCTG CTCTGGTTCC AAATAACGGC	1560
CCATGAGAAT TTGGCTGGCT CCAAAGAGTT CTTGTAGGGC GATAACAGAG TAGAGGAGAC	1620
TGGTATCCTT AATCACGTA ACAAACTGAG AAATGATGGC TGGTAGCATT TTGCGGATGG	1680
CTTGTGGGAG AATGATGTAG TAGAGGATTT GGGCTGAGGT GAAGCCTTGT GACATTCCTG	1740
CTTCGTACTG TCCCTTGTCT ACGGCATTGA GACCGCCTCG AATAATCTCA GCCAAGGCTG	1800
CTGATGTAAG GAGAGTAAAG GCTGTAATAC CTGCTGGTGT GGATTTTCAAT TTGAACACCA	1860
AAAAGATAGT AAAAATCCAG AGAAGGTTGG GAACGTTCCG CACAAACTCG ATATAAATAC	1920
TGGAAATAAT CCGTAAGACA GCATTTTTCG CATTCTCTGT GACAGCTAGC ACCGTACCGA	1980
TGATAGTAGA GAGGATGATG GCAATCAGAG AAATATAGAG GGTCAAGCCA AATCCTTTAA	2040
AGATAAAGAC TAGGTTATCT GGGGTTAAAA CTTCTAAAAT AGATTCCATA GTAACCTCCT	2100
AAAGTGAATA GGCTTTTTTG TTGGCTTGCT CCATCTTGCG ACCAAACTGG GCAACAGGGA	2160
ACCATAGAGC AAAGTAGAGA AGAGCAGCAC CTAAAAAGGC TGGTATATAG TTTCCGTTGA	2220
GAGCCGACCA AGACTTAGTC ACAAAATCA AGTCTACTCC AGAGATGATA GCTACAGTAG	2280
AGGTCTTCTT GATGAGGTTA ACAATTTGGT TGGTCAATGG AGGGAGAATG ATGCGGAAGG	2340
CCTGAGGCAA GATAATCAAG CGCATGGCAC TGATATAGGT AAAACCTTGC GACAAGGCCG	2400
CCTCCATCTG ACCACTAGGA ATAGACTGAA TCCCTGAACG AATAACCTCA GCGATATAAG	2460
CGCCGTGATA GAGTCCACG CAGAGAACGG CTGTCCAATA AATTGGAATC ATGATGATAT	2520

468

GGTCACTGAT AAGAGGTAGG CCATAAAAA CAATAACAAA CTGCACCAAG AGGGGAGTAT	2580
TTTGGTAAAA TTCAACAAAG ATGCGAGCTA AAATGCGTAA AATTGGACGT TTA CTGGTTG	2640
ACATGGCACC AAAGAAGATG CCCAAAACCA TAGCGAGGAT AAAGGAACCA ACCGCTAGGG	2700
CAAGGGTGAA GAGGAAACCA TTGAAAAATT GTCCAAAATC CTGAAAATAG GCTGTCCAAG	2760
ATGATAAATC TGTCATGGGG TGTCCTCCTT AATCTGCAGT ATGGCTAGAT GGT TGTAGCT	2820
TGTAACGGTC ATAAAGTTTC TGCAAACTAC CATCTTGCT CCATTTAGTA ACCAAGTTAT	2880
CAAGATAGTC GTTGAGCTCT GTATTTGATT TCTTGGTAAC AATACCGTAG TCAGATGGCT	2940
TGAAACTATC ATCTAGTAGT GCTGTCCGTT TACTAGTGTA GCCAGATAGA ATAGAGCGGT	3000
CAACGGAAAA GGTATCGATA CGATGAGCGT GCAGGGAACT AATCAATTCT GGGTAGGAAC	3060
CAAGTTCGAC GAATTTAAAC TTCAGACCTT TCTTTTACC CAGTTCAGTA ATCAGGCGTT	3120
GGGTGATAGA ACCTTGGGCG ACTCCGATGG TTTTGGCGTT TAGGTCTCTA ATCTTTTGA	3180
TTTTGGCAGA TTTATTGACC AAAAATCCAG AAGCGTCTGT GTAGTAGGGA CTGGTAAAGT	3240
TGTAGAGTTT TTTGCGTTCC TCCGTGATGG TAAAGGTCGC GATATCCATA TCGACCTGT	3300
CATTGTCTAG AAGGGGGCCG CGGGTTTGTG CTGTAACCGG CACATAGCGA ATCTTGACCT	3360
TGAGTTCATC AGCTACCATC TTGGCCAAGT CGGTTTCGAT ACCAGAATAA GTACCGGTCT	3420
TGGGATCTTT GTAACCAAAA TTGGGAACGT CTGTGTTGAC ACCGACAACC AGTTCCGCTC	3480
TTTTTTGAAT GTCTGCGATA CTGTGATCAG CCTGGACTGG TTTGGCAGCA GCAAGGCCGA	3540
AAAGGCTAAT CAATAATGCT GATAAAAAGA ATTTTTTTC ATAGGCGCCT CCTTATTGA	3600
CTTTGTCACT TCGTGGTTG ATAATTTTGC TGAGGAATTG TTGGGCACGA GGTTCGCTTG	3660
GATTGTCAAA AAAGTTATCG ACATCTGTCG TATCTACTAA AACTTCTCCG TCGGCCATAA	3720
AGATAATGCG GTCCGCAACC TCTCGAGCAA AGCCCATTTT GTGGGTAACG ATGATCATGT	3780
TCATCCCATC ATGCGCCAGT TTCTGCATAA CTGCTAGAAC ATCTCCGATA GTCTCAGGAT	3840
CAAGAGCAGA TGTTGGTTCA TCAAAGAGGA GGAGTCCCG ATGCATAGCA AGACCACGAG	3900
CGATGGCGAT CCGCTGTTTT TGTCACCAG ATAGCATGGC GGGATAGGAA TCTTTCTGT	3960
CCCACATATT TACAAATTCC AGATATTTTT GGGCGSTTTT TTCAGCTTCT TTTTATCAA	4020
TTCTAGAAC TTCAATGGGT GCAAGCGTTA CGTTTTCTAA CACAGCTTG TGTGGATAAA	4080
GGTTAAATG TTGAAAAACC ATGCCGACTT CCTTGCGAAG AGGTACCAAA TCTTTCTGGC	4140
TGGCACCAGC AACTTGGTGC CCATTGACTA GGAGACTTCC TTTGTCAACA GTCTCTAAAC	4200
CATTGATCGT ACGGATAAGA GTGGACTTCC CAGAGCCAGA AGGTCCAAGC AGGACAACAA	4260
CTTGTCTTTT TTCAAAACCG AGATTGATGT TCGGGAATGC GTGGTAGTCT CCGTAATATT	4320

469

TTTCGACGTT	TTTAAATTCT	ACTAAAGCCA	TGAGAGATCT	CTATTGTGTT	ATATTTTATA	4380
ACACGGTTCT	ACAATAAAAG	AATGTTCTTG	TCAAATCATA	TCTGAAAAAA	TTCAGTATAG	4440
TGAAATAAGA	ACAGGAAAAA	TCGATCGGGA	CAGTCAAATC	GATTTCTAAC	AATATTTTAG	4500
AAGTAGAGGT	GTAATATTCT	AGTTTCAATA	TACTATAAAA	TGTTATAAAA	AAGCAATCTG	4560
GATAGAGAAA	ACGTCTAAAT	CATGTTATAA	TGAAGCAATA	GAATTCTTAG	AAAGAGTGGG	4620
TGTCTTTTGG	ATAACACCTA	CTTATGAATG	GCAGTTTGCC	CTGCAGGTAG	AAGATGCGGA	4680
TTTTACAAAG	ATAGCCAAGA	AGGCTGGACT	GGGCTCTGAG	GTGGCTCGGT	TATTGTTTGA	4740
GAGAGGGATT	CAGAACCAAG	AAAGTCTGAA	GAAGTTTFTA	GAACCTTCCT	TGGAGGACTT	4800
ACATGATGCT	TATCTGCTCC	ATGATATGGA	CAAGGCAGTG	GAGCGGATTC	GTCAGGCTAT	4860
TGAAGAAGGG	GAAAATATTC	TTGTTTATGG	AGACTATGAT	GCGGATGGCA	TGACTTCGGC	4920
TTCTATTGTG	AAGGAAAGTT	TGGAACAAC	TGGTCTGAG	TGCCGAGTTT	ACCTGCCAAA	4980
TCGTTTACC	GATGGCTATG	GCCCTAATGC	TACTGTTTAT	AAATACTTTA	TCGAGCAAGA	5040
AGGGATTTC	TTGATTGTGA	CGGTGGACAA	TGGGCTTGCT	GTCATGAGG	CTATTGCATT	5100
GGCTCAGTCT	ATGGGAGTAG	ATGTCATTGT	GACAGACCAT	CATTCCATGC	CTGAAACCC	5160
GCCAGATGCT	TATGCTATTG	TCCATCCTGA	ACATCCAGAT	GCGGATTATC	CTTTTAAATA	5220
TTTGGCTGGT	TGTGGAGTTG	CTTTCAAGTT	GGCTTGTGCC	CTGTTAGAAG	AAGTGCAAGT	5280
GGAATTGCTT	GATTTGGTCC	CTATTGGAAC	TATTGCAGAT	ATGGTGAGTC	TGACGGATGA	5340
AAATCGTATC	TTAGTTCAAT	ATGGTCTGGA	AATGTTGGGT	CATACCCAGC	GCATTGGTCT	5400
GCAAGAAATG	CTGGACATGG	CTGGGATTGC	TGCCAACGAA	GTAACAGAAG	AAACGGTTGG	5460
TTTCCAGATT	GCTCCTCGTT	TGAATGCCTT	GGGTCGCTTG	GATGATCCCA	ATCCTGCCAT	5520
TGATTGTTG	ACTGGATTTG	ATGATGAGGA	AGCGCATGAG	ATTGCCCTTA	TGATTACCA	5580
GAAAAACGAA	GAGCGCAAGG	AAATCGTTCA	GTCTATCTAT	GAAGAAGCCA	AGACCATCGT	5640
GGATCCTGAG	AAGAAGGTTT	AGGTCTTGGC	CAAGGAAGCC	TGGAATCCTG	GGGTTCTAGG	5700
AATCGTGGCT	GGTCGTTTAT	TGGAAGAATT	GGGACAGACA	GTCATTGTTT	TTAATATAGA	5760
AGACGGTCGT	GCCAAGGGCA	GTGCTCGTAG	TGTGGAAGCG	GTCGATATTT	TTGAAGCTCT	5820
GGATCCCCAT	CGAGACCTCT	TCATCGCCTT	TGGAGGTCAT	GCAGGTGCAG	CGGTATGAC	5880
GCTGGAAGTT	GAGCAACTCT	CAGATTTATC	TCAGGTTTTG	GAAGATTATG	TTCTGAAAA	5940
ACGTGCAGAT	GCTGGTGGCA	AGAATAAGTT	AAACCTAGAT	GAAGAGTTGG	ATTTGGAGGC	6000
ACTTAGCTTG	GAAACGGTCA	AAAGTTTGA	ACGTTTAGCT	CCTTTTGGAA	TGGATAATCA	6060

470

GAAACCTATT TTTTATATCA AGAATTTTCA GGTCGAAAGT GCTCGTACTA TGGGGGCAGG	6120
TAATGCCCAT CTAAAGCTGA AAATTTCCAA GGGTGAGGCG AGTTTGAAG TGGTAGCCTT	6180
TGGTCAAGGC AGATGGGCGA CAGAGTTTTC TCAAACCAAG AATCTAGAGT TAGCGGTAA	6240
ATTGTCTGTC AACCAATGGA ATGGCCAAAC TGCCCTCCAG TTGATGATGG TGGATGCGCG	6300
AGTGGAAGGT GTTCAACTTT TTAACATTCTG TGGAAAAAAT GCAGTCTTGC CAGAAGGTGT	6360
TCCAGTCTTG GATTTTCCTG GAGAACTGCC AAATCTTGCG GCTAGTGAAG CTGTTGTCGT	6420
AAAAACATT CCAGAGGATA TTAATCAGCT GAAGACCATT TTTCAGGAAC AGCATTTCTC	6480
TGCTGTCTAT TTCAAAATG ATATTGACAA GGCTTATTAT CTGACAGGTT ATGGGACTAG	6540
AGATCAGTTT GCCAAATTGT ACAAGACTAT TTACCAGTTC CCAGAGTTTG ATATTGCTA	6600
CAAGCTGAAA GATTTGGCTG CATATCTTAA TATTCAACAA ATCTTGCTGG TCAAGATGAT	6660
TCAAGTATTT GAAGAACTAG GCTTTGTGAC GATAAAAGAT GGTGTGATGA CAGTCAATAA	6720
AGAGGCGCCA AAGCGGGAGA TAGGAGAAAG TCAAATTTAC CAAAATCTCA AACAAACCGT	6780
TAAAGACCAA GAAATGATCG CGCTGGGTAC GGTGCAAGAA ATTTATGATT TTTTGATGGA	6840
AAAAGAGTAG AAGTTAGGAA AGAGTTGGGA AATCAACTCT TTTTGAATAA CAGACCTTCA	6900
TTTTGAAAAT CATCAAAAAA ATCGTATAAT GGTAGGAAAA GATTCGGCTG AAAGTATCAG	6960
AACCTTTTGA ATAAGAGGGT AGAATTGCCC TATAATCAAG ATAAACTAAG ATTTTGAGG	7020
AAAAATGAGT AATATCAGTT TAACAACACT TGGTGGTGTG CGTGAGAATG GAAAAATAT	7080
GTACATTGCT GAAATTGGAG AGTCCATTTT TGTTTTGAAT GTAGGGTTAA AATATCCTGA	7140
AAATGAACAA TTAGGGGTCT ATGTGGTGAT TCCAAACATG GATTACCTTT TTGAAAATAG	7200
CGACCGTATT GCTGGGGTTT TCTTGACCCA CGGGCATGCG GATGCCATTG GTGCTCTACC	7260
GTATCTCTTG GCAGAGGCTA AAGTTCTGT ATTTGGGTCT GAGTTGACCA TTGAGTTGGC	7320
AAAGCTCTTT GTCAAAGGAA ATGATGCCGT TAAGAAATTT AATGATTTC ATGTCATTGA	7380
TGAGAATACG GAGATTGATT TTGGTGGGAC AGTGCTTTCC TTCTTCCCTA CGACTTACTC	7440
CGTTCCAGAG AGTCTGGGAA TTGTCTTGAA GACATCGGAA GGAAGCATCG TTTATACAGG	7500
TGACTTCAAA TTTGACCAAA CGGCTAGTGA ATCTTATGCA ACTGATTTTG CTCGTTTGGC	7560
AGAGATTGGT CGTGACGGCG TCCTGGCTCT CCTCAGTGAT TCGGCCAATG CAGACAGCAA	7620
TATTCAGGTG GCTAGTGAAT GTGAAGTTAG GGATGAAAT ACCCAAATA TTGCTGACTG	7680
GGAAAGTCTG ATCATCGTTG CAGCTGTTTC CAGTAATCTT TCTCGTATTC AGCAGATTCT	7740
TGACGCTGCG GATAAACAG GTCGACGTAT CGTCTTGACA GGATTTGATA TTGAAAATAT	7800
CGTCCGCACA GCGATTCTG TTAAGAAGTT GTCTTTAGCC AACGAAATTC TTTTGATTAA	7860

471

GCCTAAAGAT ATGTCTCGCT TTGAAGACCA TGAGTTGATT ATTCTTGAGA CAGGTCGTAT 7920  
GGGTGAGCCT ATCAATGGAC TTCGTAAGAT GTCGATTGGT CGCCATCGTT ATGTAGAAAT 7980  
CAAGGATGGG GACCTAGTCT ATATTGCTAC GGCTCCGTCT ATTGCTAAAG AAGCCTTTGT 8040  
TGCGCGTGTG GAAAATATGA TTTATCAGGC AGGTGGGGTT GTCAAATGA TTACCCAAAG 8100  
TTTACATGTA TCAGGGCAGC GAAATGTGCG TGATTTGCAG CTGATGATCA ATCTTTTGCA 8160  
ACCTAAGTAC CTCTTCCTG TCCAAGGGGA GTATCGTGAG TTGGATGCTC ACGCTAAGGC 8220  
TGCCATGGCA GTTGGGATGT TGCCAGAACG CATCTTCATT CCTAAAAAGG GGACGACCAT 8280  
GGCTTACCAG AATGGAGACT TTGTTCCAGC TGGATCGGTT TCAGCAGGAG ATATCTTGAT 8340  
TGATGGGAAT GCCATTGGTG ATGTTGAAA TGTGTTCTT CGTGACCGTA AGGTCTTGTC 8400  
AGAGGATGGA ATTTTCATCG TGGCTATTAC AGTCAACCGT CGTGAGAAGA AAATTGTGGC 8460  
TAGGGCTCGT GTTCACACGC GTGGATTTGT TTATCTCAAG AAGAGTCGCG ATATTCTCCG 8520  
TGAAAGTCA GAATTGATTA ACCAAACGGT AGAAGAGTAT CTTCAAGGAG ATGACTTTGA 8580  
CTGGGCAGAT CTCAAAGGTA AGGTTCGTGA CAATCTGACC AAGTACCTCT TTGATCAAAC 8640  
CAAGCGTCCG CCAGCCATTT TACCAGTAGT CATGGAAGCA AAATAATCGT TGAAATAAAC 8700  
AGAGAGAAAG TCGAGTTTCG GCTTTTCTT ATAGAAAAAT AGAAGGAGAA AATCATGGCA 8760  
GTGATGAAAA TCGAGTATTA CTCACAAGTA TTGGATATGG AGTGGGGGGT GAATGTCCTC 8820  
TACCCTGATG CCAATCGAGT GGAAGAACCA GAGTGGAAG ATATTCCCGT CTTGTACCTT 8880  
TTGCACGGGA TGTCTGAAA TCATAATACT TGGCTTAAGC GGACCAATGT AGAACGCTTG 3940  
CTTCGAGGAA CTAATCTCAT CGTTGTTATG CCCAATACCA GCAATGGTTG GTACACCGAT 9000  
ACCCAGTATG GTTTTGAATA CTACACGGCT CTAGCAGAGG AATTGCCACA GGTTCGAAA 9060  
CGCTTCTTCC CTAATATGAC GAGCAAGCGT GAAAAGACCT TTATCGCTGG TCTTTCTATG 9120  
GGAGGCTACG GCTGCTTCAA ACTGGCTCTT ACGACAAATC GTTTTCTCA TGCAGCTAGT 9180  
TTTTCAGGTG CCCTCAGCTT TCAAACTTT TCTCCTGAAA GTCAAAATCT GGGAACTCCA 9240  
GCCTACTGGA GAGGTGTTTT TGGAGAGATT AGAGACTGGA CAACTAGTCC CTATTCTCTT 9300  
GAAAGTCTGG CTAAAAATC GGATAAAAAG ACCAACTTT GGGCGTGGTG TGGCGAACAG 9360  
GATTTCTTGT ACGAAGCCAA TAATCTCGCA GTGAAAAATC TCAAAAAACT AGGTTTTGAT 9420  
GTGACCTATA GCCATAGCGC TGGAACTCAC GAGTGGTACT ACTGGGAAAA ACAATTGGAA 9480  
GTTTTTTTAA CAACCTACC AATTGATTTT AAATTAGAAG AGAGACTGAC TTAGTTTGAA 9540  
CTTCAGCATA GGGGGAGTAG AACTAAAATA AAATATCTTT TCACTAGACT TTTCAAACGm 9600

472

AAGTAGTAGA ATAGTAATAA AATACTGGAG GAAAGAGAGT AGGAAATGTA CCGTTATCAA	9660
ATTGGCATTCCACATTAGA ATATGATCAG TTTGTCAAAG AACATGAATT AGCCAATGTA	9720
TTACAAAGTA GTGCTTGGGA GGAAGTTAAG TCTAATTGGC AACATGAGAA GTTTGGTGTT	9780
TACAGGGAAG AAAAATTACT GGCGACAGCT AGTATTTTGA TTAGAACTCT TCCGCTAGGC	9840
TATAAAATGT TTTACATCCC AAGAGGACCT ATATTGGATT ATGGGGATAA AGAACTCTTG	9900
AATTTTGCCA TTCAGTCTAT TAAGTCCTAT GCTCGCAGTA AGAGAGCGGT TTTTGTGACT	9960
TTTGACCCAA GTATTTCCT ATCTCAAAGT TTAATCAATC AGGAAAAGAC AGAATTTCTT	10020
GAAAACTCTGG CTATTATTGA TAGTTTGCAA CAAATGGGAG TAAGGTGGTC AGGAAAAACG	10080
GAGGAAATGG GAGACACCAT TCAACCTCGT ATTCAGGCGA AAATATACAA GGAAAAATTT	10140
GAAGAAGATA AACTTTCCAA GTCAACAAAA CAGGCTATTC GAACAGCAGC AAACAAAGGG	10200
CTTGAGATTTC AATATGGTGG ACTGGAACATA TTAGATTCAT TTTCCGAGTT GATGAAAAAA	10260
ACTGAGAAGC GAAAAGAGAT TCATTTGAGG AATGAAGCCT ATTATAAAAA ATTGTTAGAT	10320
AATTTTAAGG ACAAGGCCTA TATCACCTTG GCCACCTTGG ATGTTTCTAA ACCTTCGCAA	10380
GAGTTAGAAG AACAGTTAGC GAAAAATAGA GCCTTGAAG AGACCTTTAC TGAGTCGACT	10440
CGAACTTCAA AAGTAGAAGC GCAGAAGAAG GAAAAAGAAC GTTTGTTAGA GGAATTGACC	10500
TTCTTGCAAG AATATATAGA TGTAGGTCAA GCGAGAGTTC CTTTAGCGGC TACTTTGAGT	10560
TTGGAATTTG GACTACCTC TGTCAATATA TATGCTGGTA TGGATGATGA TTTTAAACGT	10620
TACAAATGCAC CAATTTTAAC TTGGTATGAA ACGGCTCGCT ATGCCCTTGA ACGAGGTATG	10680
ATCTGGCAAA ATTTAGGTGG TGTGAAAAC TCTCTCAATG GTGGACTTTA TCATTTTAAG	10740
GAAAAATTTA ATCCAACGAT TGAAGAATAC TTGGGTGAAT TTACAATGCC CACTCATCCT	10800
CTCTATCCTC TGTTAAGACT TGCTCTTGAT TTCCGTAAAA CATTAAAGAA AAAACATAGA	10860
AAGTAAGTAT ATGGCACTAA CAACACTCAC GAAAGAAGAG TTTCAGACTT ATTCTGATCA	10920
GGTTTCTTCT CGTTCTTTA TGCAATCTGT CCAGATGGGG GATTGCTAG AAAAAAGAGG	10980
GGCTCGAATT GTTTATCTTG CTTTGAAACA AGAAGGAGAA ATTCAAGTTG CAGCTCTGGT	11040
TTATAGCCTG CCCATGCTGG GTGGTCTGCA TATGGAAGTC AATTCGGGGC CGATTTATAC	11100
CCAACAAGAT GCTCTTCCAG TTTTATATGC AGAGTTAAAA GAATATGCCA AGCAAAATGG	11160
TGTATTAGAG TTGCTTGTA AACCCTATGA AACTTATCAA ACTTTTGATA GCCAAGGTAA	11220
TCCAATAGAT GCTGAGAAAA AAAGTATTAT TCAAGATTTG ACTGATTTAG GTTATCAATT	11280
TGATGGCTTA ACAACAGGTT ACCCAGGTGG AGAACCAGAT TGGTTTACT ATAAAGATTT	11340
AACTGAATTA ACTGAAAAGA GTTTGCTTAA AAGTTTTAGC AAAAAAGGTA AACCCTTGGT	11400

473

GAAAAAGGCT GAAACCTTTG GCATTGGTT GAAAAAGTTA AAACGTGAAG AACTATCGAT	11460
TTTTAAGAAT ATAACAAAAG AAACCTCTGA ACGTAGAGAA TATAGTGATA AAAGTTTAGA	11520
ATATTATGAG CATTTTATG ATACTTTTGG AGAACAAGCG GAGTTTCTCA TAGCAAGCTT	11580
AAATTTTTCG GACTATATGA GCAAATTGCA AGGTGAACAA ACTAACTAG AAGAAAACCT	11640
GGACAAGTTG CGACTTGATT TGAGTAAAAA TCCTCATTCT GAGAAAAAAC AAAATCAACT	11700
GAGAGAATAT TCTAGTCAAT TTGAAACGTT TGAAGTTCGA AAAGCAGAAG CGCGAGACTT	11760
GATTGAAAAA TATGGAGAAG AAGATATTGT TTAGCTGGG AGTTTATTTG TTTATATGCC	11820
TCAGGAAACG ACTTATCTCT TTAGTGGTTC CTACACTGAG TTAATAAGT TCTATGCCCC	11880
TGCACTGCTT CAAAAATATG TTATGTTGGA AAGCATAAAA CGTGGGAATAC CTAAATACAA	11940
CTTCCTAGGC ATTCAAGGGA TTTTGTATGG AAGTGATGGT GTTTTGCCTT TTAACAGAA	12000
TTTAAATGGC TATATTGTAC GCAAAGCAGG TACTTTCCGT TACCATCCAT CGCCTTTAAA	12060
ATACAAAGCT ATCCAGTTAC TCAAAAAAAT AGTAGGACGT TAAGATGAAA AAGTCAGTAT	12120
TTAGATTTCT TTAGCTTCT TTAGTAAAA TAATTCTTAT TTGCTAGAAA GGTGGAGAGA	12180
CATGCGCTGG CTTTTCGTT TGATAGGGGC TTTCTTTTCT TTTGTGTGGC GTTTGTTTTG	12240
CGGTCTGGTT TGGATAGTTG TGCTCTTATG TGTGCTTGCT TTCGGACTTC TCTGGTATCT	12300
GAACGGAGAT TTTCAAGGAG CGCTAAAGCA AGCAGAACGG TCAGTAAAAA TTGGTCAACA	12360
AAGTATTGAC CAATGGGAGA AAACAGGGCA ACTGCCTAAG TTAAGCCAGA CAGATAGTCA	12420
CCAGCATTCT GAAGGAAGGT GGGCACAGGC CTCTGCTCGT ATTTACCTGG ATCCGCAGAT	12480
GGATTCACGC TTTCAAGAGG CTTATTAGA AGCAATCCAG AACTGGAATC AAAGTGGTGC	12540
TTTAACTTT GAACTCGTGA CTGAGTCTAG TAAGGCGGAT ATTACGGCTA CGGAGATGAA	12600
CGACGGAGGC ACTCCTGTGG CAGGAGAGGC GGAAAGTCAA ACTAATCTCT TAACAGGGCA	12660
ATCTCTGTCC GTAACGGTGC GGTGGAATCA TTATTATTG TCCAATCCAT ACTATGGCTA	12720
CTCCTATGAA CGCCTTGTCC ATACGGCAGA ACATGAGTTA GGTCAATGCGA TTGGCTTGGA	12780
CCATACAGAT GAGAAGTCTG TCATGCAACC AGCAGGTTCC TTTTATGGTA TCCAGGAAGA	12840
GGATGTTGCA AACCTCCGAA AAATATATGA GACTAGTGAG TAGGGTACTA TCTTTCCCTA	12900
CTTTTTTTCG TATAATGGAA CTATGAACAA CTTGATTAAA TCAAACTAG AGCTCTTGCC	12960
GACCAGCCCT GGTGCTACA TTCATAAGGA TAAAAATGGC ACCATTATCT ATGTAGGAAA	13020
GGCTAAAAAT CTGCGTAATC GAGTACGGTC CTATTTTCGT GGAAGTCATG ATACCAAGAC	13080
AGAGGCTCTG GTGTCTGAAA TTCTGGATTT TGAATTTATT GTTACGGAGT CTAATATTGA	13140

474

GGCACTTCTC CTAGAAATCA ACCTGATCAA GGAACAAG CCAGTACA ATATCATGCT	13200
CAAGGATGAC AAGTCCTATC CTTTCATCAA AATCACCAT GAGCGCTATC CACGCTTGAT	13260
TATCACTCGT CAGGTCAAAA AGGACGGAGG TCTTTATTTT GGACCCTATC CCGATGTGGG	13320
GGCAGCCAAT GAAATCAAGC GGTGCTGGA TCGGATATTC CCTTTTCGTA AGTGTACCAA	13380
CCCCCCTCT AAGGTCTGTT TTTATTACCA TATCGGCCAG TGTATGGCCC ACACCATCTG	13440
TAAGAAGGAT GAGGCTTATT TCAAGTCTAT GGCCAGGAG GTGTCTGATT TTCTGAAAGG	13500
TCAGGATGAC AAAATCATCG ATGATCTCAA GAGTAAATG GCAGTAGCAG CACAAAGTAT	13560
GGAGTTTGAA CGTGGGGCGG AATACCGTGA CCTGATTCAG GCTATTGGAA CGCTTCGAAC	13620
CAAGCAACGG GTCATGGCGA AAGATTGCA AAATCGCGAT GTCTTTGGCT ACTATGTGGA	13680
TAAGGGCTGG ATGTGTGTGC AGGTTTCTT TGTCCGTCAG GAAGCTCAT CGAGCGCGAT	13740
GTCAATCTCT TCCCCTACTT CAATGATCCA GATGAGGATT TTTTGACCTA TGTAGGACAA	13800
TTCTATCAAG AAAAATCTCA TCTAGTTCCC AATGAGGTAC TGATTCCGCA GATATTGACG	13860
AAGAAGCTGT CAAGGCTTTG GTGGATTCCA AGATTCTTAA GCCTCAACGT GGAGAGAAAA	13920
AACAACTGGT CAATCTAGCC ATAAAAATG CTCGTGTTAG TCTAGAGCAG AAGTTCAATC	13980
TGCTAGAAAA ATCTGTCGAA AAGACTCAAG GAGCTATTGA AAATCTAGGG CGTTTGCTCC	14040
AAATCCCGAC CCCAGTACGT ATCGAGTCCT TCGATAACTC TAATATCATG GGAAGTAGCC	14100
CTGTTTCGGC TATGGTGGTC TTTGTCAACG GTAAACCGAG TAAGAAGGAT TACCGTAAGT	14160
ACAAGATAAA AACGGTTGTT GGACCAGACG ACTATGCCAG CATGAGAGAG GTCATTGCGA	14220
GACGCTATGG TCGAGTACAG CGTGAGGCTT TGAATCTCC AGATTTGATT GTGATTGATC	14280
GGGGGCAAGG TCAAGTCAAT ATCGCTAAGC AGGTTATCCA AGAGGAACTG GGCTTGATA	14340
TTCCAAATGC TGGGCTGCAA AAGAATGATA AGCACCAAAC CCATGAATTG CTCTTTGGAG	14400
ATCCGCTTGA GGTGGTGGAT TTGTCTCGCA ATTCTCAGGA ATTTTCTCTC CTCCAACGCA	14460
TCCAAGATGA GGTGCACCGC TTTGCTATCA CTTTCCACCG CCAACTGCGC TCCAAAAATT	14520
CTTTCTCATC TCAATTGGAT GGGATTGACG GTCTGGGACC TAAACGCAAG CAGAATCTTA	14580
TGAAGCATTT CAAGTCTTTG ACCAAAATCA AGGAAGCCAG TGTGGATGAG ATTGTCCAAG	14640
TTGGGGTACC TAGAGTCGTT GCAGAGGCTG TGCAAAGAAA GTTGAACCCG CAGGGAGAAG	14700
CCTTGCCCTCA AGTAGCAGAA GAAAGAGTAG ATTACCAAAC GGAAGGAAAC CACAATGAAC	14760
CATAAAATCG CAATTTTATC AGATGTTTAT GGCAATGCGA CGGCGCTAGA AGCAGTGATT	14820
GCAGATGCTA AAAATCAAGG GGCCAGTGAA TATTGGCTTC TGGGAGATAT TTTTCTTCTC	14880
GGTCCAGGCG CAAATGACTT AGTCGCCCTG CTAAAGGACC TTCCTATCAC AGCAAGTGTT	14940



475

CGAGGCAATT GGGATGATCG TGTCTTGAG GCTTTAGATG GGCAATATGG CTTAGAAGAC	15000
CCACAGGAAG TTCAGCTCTT GCGTATGACA CAGTATTTGA TGGAGCGAAT GGATCCTGCA	15060
ACGATTGTCT GGCTACGAAG CTTGCCTTTG CTGGAAAAGA AAGAAATTGA CGGATTGCGC	15120
TTTTCTATCT CTCATAATTT ACCTGACAAA AACTATGGTG GTGACTTGCT AGTTGAGAAT	15180
GATACAGAGA AATTTGACCA ACTGCTAGAT GCGGAAACGG ACGTGGCAGT TTATGGTCAT	15240
GTTCACAAGC AGTTGCTTCG TTATGGAAGT CAAGGGCAAC AAATCATCAA TCCAGGGTCG	15300
ATTGGCATGC CCTATTTTAA TTGGGAGGCG TTAATAAATC ACCGTTCCCA GTATGCCGTG	15360
ATAGAAGTTG AAGATGGGGA ATTACTCAAT ATCCAATTTT GTAAAGTTGC TTATGATTAC	15420
GAAGCTGAGT TAGAATTGCC CAAGTCCAAG GGGCTTCCCT TTATCGAAAT GTATGAAGAA	15480
CTGCGTCGTG ACGATAACTA TCAGGGGCAC AATCTGGAAT TATTAGCCAG CTTAATAGAA	15540
AAGCATGGGT ATGTAGAGGA TGTGAAGAAT TTTTGTGATT TTTTGTGAAGA GTTTCCTAAA	15600
ATAGCCAATG CAAACTAAAA AAGCGATTTG CTGGTCCAAT CGCTTTTAGT ATATCTTATA	15660
CTCAATGAAA ATCAAAGAGC AAAC TAGGAA GCTAGCCGTA GGTGCTCAA AGCAGAGCTT	15720
TGAGGTTGCA GATAAAGCTG ACGTGGTTTG AAGAGATTTT CGAAGACTGT TATTGTAACT	15780
GAGATTGATC TGGGAGGTAA GAACCACTA GATAGGTATT GCTGAUTTTT TCAAGGGTTC	15840
CGTCTTGATA GAGTTCTTTG AGCGCTTTAT CAAATTGCTC TTTAACTCT TTTTGGTCGC	15900
TTGAGAAAAT GATATAATTG CTGGGGCTAT CTGCAGAAGG TAAATCAACG ACTGAGAGGT	15960
CTAAACCACG GTCCTTGATA ATCTTTTGAA CGGATACCTT CTCAAAACT AGGAAATCAA	16020
ACTCTCCGTT AGCAAGGTCT AGGATTCGTT TACCAATATC CTCACCAGAA AAATTAATTG	16080
TAGCGGGATT ATCAGTGTGT TTCTGATTCC AGTTATTGAT GAATTGAGCG TTAGAAGTTC	16140
CGGTATCCTC TTGTGTTGTT TTACCAGCGA TCTGGTCAAG AGAAGTCAA GGATTTTCT	16200
TGTTGCTGAC AAGGACGAGG GGATTGTTGG AAATTGGAAG CGAGTAAAGG TATTTTTCAG	16260
CACGCTCTTT TGTGTAATC AAGTTATTGG CCGCAGCCTG ATAGTGACCA GAATCAAGTC	16320
CTGGGAAGAT GCTCTCCAG GCGGTTCTTT GGAATTGAAT CTCGTAGTCG CTGAGTTTTT	16380
CATCTACTGC CTTTAAAACT TCGATATCAA AGCCTGTCAG ATTGCCCTTG TCTTCGTAGT	16440
CAATGGTGG CACGTCGCCA GCTGTAGCAA GGACGATTGT CTTTGTAGCG CTAGTCTCTT	16500
TGGGTGTAGC TTGATTCTCA CAGGCAACCA AAAATGGTAG GATAGCTAGT AATAGGCTAA	16560
ATTTTTTCAT ACTGTCTCCA TTCAAATGTA AAG	16593

(2) INFORMATION FOR SEQ ID NO: 53:

476

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3510 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

GGGATATCCT TATATCCTTG TTCCTGGAAC CATTGTGGGA ATTGCTCAAC AGTTTTTTTCA	60
CCTTGAATTC CTGGTGCAAT GACAGTAAGA ATTTGGAAT CACGATCTGG TTTCGCCGCT	120
AGTTCCATCA ACTCTGGCAT ACTTTTCTTG CATGGACCAC ACCATGAAGC CCAAAACTTC	180
AAGTAAACCT TTTTACCCTT AAAATCAGAT AACTTAACCT CTTTGCCATC CATGGATTGC	240
AATGTGAAGT CTGGAGCATC TTTTCCAACA GCAATTTGTT GTACAGTCGT TTGTTGTTTT	300
GGCTGTTGTG CTGCTTGAGT CTTTTTAGTT TCTTCCTCAC CACAGGCCAT CAATACAACT	360
AATGACAAGA GACTTAAGCC AGCAAACATT ACTTTTTTCA TTGTCTCTCC TTTATTCAAA	420
AATTCCAGCT AGAACATTTA CTTGTCCTAA TAGTAACAAA ATTCCCATT AACAATGAG	480
GAAACCACCA ATTTTCTTTA GTAGCATCAT ATGACGCTTG ATTTTACTAA AATATGGCAT	540
GACTAGACCT GAAGCTAGTG CCAATACCAA GAAAGGAAGG GCCATGCCAG ACTGTAAATG	600
AGAGTATAAA TCGCTCCTTG CCAAGCGCCA TTGCCTCCAG AAGCCGCAAG TGCTAAAACA	660
GAACTTAAAA CTGGACCAAT ACAAGGTGTC CAACCAAAGC TAAAGGTAAT ACCAAGTAAA	720
AAAGCTGACC AATAACGATT AGAATCTGAT TTTTAAAGG TAAAACTTT TTGAACCTCT	780
AATTTCTTCA AATGAAAAAT TTCCATCTGG TGAAGACCCA AAATGATAAT AATAGCTCCC	840
ATGCCATATC GAAACCAATT TGCATAGAGA ATATGACCAA AGTAACCAGC ACCAAAGCCT	900
AGAATAAAGA AAATGAGAGA GATACCAGCG ATAAAGCAAA GTGTTTGAAT CAAGCCTGAC	960
CAGAGAACCT TTCTCCCAA CAAAGAAAAG CTTTTTGAC TTTCTTGATC ATCCAATAAA	1020
ATCCCAGCAT AGACTGGCAG AAGAGGAAAA ATACAAGGAG AAAAAAGGA TAAACACCT	1080
GCTAGAAAAA CAGAGATTAA AAATACTATC GTTTCCAATA AAGAACCAAC TTTCTTAATA	1140
ATTCTAATCC TATTTTACTA TATTCAATTT TATTTGTAAG CTTTCTGCTA CGCAAAATCG	1200
TATCGGGCAC TATTGGACCA ATCTTTTCTT TTGCTAGTCA AGGCGGATCT TATCCCCCAA	1260
AATAGCCAAA AAGCAACGAC AAGGATTACT CATCGCTGCT TTTGTGAACG AAAATGTCTT	1320
TTAGGTCTGA CATTTCAATA ATCATGTTTT ACTTGAGTTT GTCAAGGATT GCTTTAAGCT	1380
CCTCTACTAG TTAGTTTCT GTCTCTGCTG AGCCATTTTC TTCTTTCACG AAATCAAGGG	1440
TTTCTTGAG AAGGTTTTGG GCTTTGGCAA GGACTTTTTT ATCCGCTTTT TCTGCATCTA	1500

477

GCTGTCCTAG AACCTTGATC AATTCCTGTC TTAATTGCTG GATT	1560
GGCGAATCAG CCAGAAGGCA ATCACGCCTA GGAGGGCAAG TAGACTGACC ACAATCACTC	1620
CTGCCGGAAC TGAGTTTGTT TCAGTCATCT TATCTGAATC CTTACTATCT TCCGTTCCTT	1680
GTTTTGCATC CTTCTTGTC TGTGCAGGCT TGCTGTCGCT AGCATTGCT TTCACATCTT	1740
TGAGAGAGTC CAAGGCAGCC CAGCCTTCAC AGACTCTACT GCAGTATGCA GACCTTACTC	1800
TGTCAAGGCA CTATCTTCGG GAGCTTTTG AGCATCTAGG AGGACAGCCT TGCTTGCATC	1860
GATTTTCGGA TCAGATACTG TTGCCAAAGC TTTCAAGCGT TGGTCTAACT CTTGACTCAA	1920
GGCACGAAGT TCAGACTTGT CAACTTGCTC TTGAGCTTGT GTGCTCGTTG AGCTAGCCGA	1980
AGCGCTTGCT ACCACTCTAG GATCTTGAGT CGGAGCTGAG CTTGGAGCTG GGACAGGGCT	2040
TGCAGGTTGA CTAGGAACAG TTATGGTATA TTGAACTAG AATAGTACAT ATGGACTTCT	2100
AAAACATTGT TAGAATTGCA TTTACTGTC CTGATCGATT TGTCTATTG TTATTTTATT	2160
TTACTATAAT AACCGATGGT GTGGTTAATG TTGGTAAGAG AAAGTTCTGA AACCAAGCTT	2220
CAAAAAAGTC GCTCGTCATC GTCTCTCGT AAGTCATTGG AGCGATTAAT TCACCATTTG	2280
TTAGACCTGC AACCAAAGAA ATCCTCTGAT ATCTTCTTCC AGATACTTTG CCTCTTATTA	2340
ACTGACCTTT TAATGAGCGA CCATATTCTC GATAAAAATA AGTATCGAAT CCTGTTTCGT	2400
CAATCTAAAC AGGTGCTAGG TGCTTTAAAC TATTAAATTT CTTAAGAAAT AAGGCTACTT	2460
TTTCTGGGTC TTGTTTCATAG TAGGTGTGGT TCTTTTTTTC GAGTGTAGCC CATAGCTTTG	2520
AGCGCATAGT GGATGGTAGT TGGATGACAG CCAAATCAG AAGCTATTTT AGTCAAATAA	2580
GCCTCTGGAT TGTGAGTAAG ATAGTTTTTA AGTCTATCTC TATCAACTTT TCTTGGTTTT	2640
GTTCCTTTTA CTTGGTGGTT TAGCTCTCCT GTTTTCTCTT TTAGCTTTAA CCAGCCATAA	2700
ATGGTATTAC GTGAGATTG GAAAACGTGT GATGCTTCTG TTATACTACC TATTCGCTCA	2760
CAATAAGAGA GAACTTTTTT ACGAAAATCT ATTGAATATG CCATAAGAAG ATTATACCAC	2820
ATTGTGTAAT ATTTTGGTT CATTTCATA TAACACAAAA TAGATTATTA TTACATAACA	2880
AAAAAGAGGT CTAAACCTCT TAACTCAATT ACTCCGCCAG TAGGACTCGA ACCTACGACA	2940
TCATGATTAA CAGTCATGCG CTACTACCAA CTGAGCTATG GCGGATTAAA GCTAAGCGAC	3000
TTCCCTATCT CACAGGGGGC AACCCCAAC TACTTCCGGC GTTCTAGGGC TTAAGTTCTG	3060
TGTTCCGCAT GGGTACAGGT GTATCTCCTA GGCTATCGTC ACTTAACTCT GAGTAATACC	3120
TACTCAAAAT TGAATATCTA TTCAATTTAA GAAAACCGTT CGCTTTCATA TTCTCAGTTA	3180
CTTTGGATAA GTCCTCGAGC TATTAGTATT AGTCCGCTAC ATGTGTCGCC ACACTTCCAC	3240

478

TTCTAACCTA TCTACCTGAT CATCTCTCAG GGCTCTTACT GATATATAAT CATGGGAAAT	3300
CTCATCTTGA GGTGGKTC CA CACTTAGATG CTTTCAGCGT TTATCCCTTC CCTACATAGC	3360
TACCCAGCGA TGCCTTTGGC AAGACAACTG GTACACCAGC GGTAAGTCCA CTCTGGTCCT	3420
CTCGTACTAG GAGCAGATCC TCTCAAATTT CCTACGCCCG CGACGGATAG GGACCGAACT	3480
GTCTCACGAC GTTCTGAACC CAGCTCGCGT	3510

## (2) INFORMATION FOR SEQ ID NO: 54:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20986 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

CGGAGAAAAA CATGGCTAAG TCAAACCTTG AAAAAGTAGA ATCAGTTGTT GGCTGGGTTC	60
GTGATAAGAA AATCACAGGC TACCGTATCT CTAAAGAAAC GAATGCGCGT GAAATGTCTA	120
TCATTGCTCT GGCGCAGGGT CGTGCAAAAG TAAAAATAT TTCATTTGAA ACAGCCCTAG	180
GCCTAATTGA TTTCTATGAA AAAAATTATG AAAAATTGA AGATTAATCT TTGGATAACG	240
GCGGATTCTT GACCTTCAAG TAGTAGAGAT AGAGAATCTG CCTTTTCATT TTGAGGACAG	300
CAAAAAGACT GCACGGTTGA TGCAGCCTTT TCTTTTATT TGAGATAGCG TTGAAGGAAC	360
TCTTTTGTTT GGTCTTCTTT AGGATTGGTG AAGAGGTCTT CTGGTTTACC TTCTTCAGCG	420
ATCACGCCCT TATCCATAAA GATAACACGG TGAGAGACAT CACGGGCAAA TTCCATTTC	480
TGGGTACGA CAATCATGGT CAAGCCTTCC TGAGCCAGGT CCTGCATGAT TTTGAGGACT	540
TCTCCAACCA TTTCTGGATC GAGAGCTGAT GTTGGTTCAT CAAAGAGAAT AGCGTCCGGA	600
TTTCATGGAGA GGGCAGGAGC GATGGCCACA CGTTGTTTTT GACCACCTGA GAGTTGTTTT	660
GGTTTGGCTT GCCAGTAGCG TTCTCCCATG CCGACCTTTT CCAGGTTTTT TTTGGCAATC	720
TTTTTCAGCTT CTGTGCGTTC GCGTTTTAGG ACAGTTGTCT GAGCGACGAT TGTGTTTTCA	780
AGAACATTGA GATTTTCAAA GAGGTTAAAG GATTGGAAAA CCATCCCCAA CTTTTCACGG	840
TATTGCGTGA GGTTCATAGC TTTTCGAGG ACGTTTTGTC CATGATAAAG GATTGTGCCA	900
TCAGTTGGTG TTCAAGTAG GTTAATGGAG CGTAGGAAGG TCGATTTTCC GCTTCCAGAG	960
CTTCCGATGA TAGAGATGAC CTCTCCCTTG TGGACAGTGA GTGAAATGTC TTTTAGCACT	1020
TCGTTTTGTC CATAGGATTT TTTGAGGTGT TTAATTTCAA GGATTGCTTG TGTCAATTAT	1080
TCAAATCCTC CGTTTGCATT TGGTTAGCAC CTGTAGTGTA GGTATCCATG TCCATTCTGC	1140

GCTCGATAAA GCGTAGGATA CGTGTACGG TGAAGGTGAG GACAAAGTAA ATCACGGCGA	1200
TGATTGTAAA TGTCTGGAAG TATTGATAGG TTTGTGTTGC CACGGTATTT CCTGAGAAAT	1260
AAAGTTCCGAC AACAGAGATA ACGTTCAATA CAGATGTATC TTTGATATTG ATGACAAATT	1320
CATTACCAGT TGCAGGTAGG ATGTTACGGA CTACCTGAGG TAGGACAATC TTACGCATGG	1380
TCTGGTTATG GGTTCATACCA AGAGCAGTCG CAGCTTCAAA TTGTCCCTTG TCAACTGCTA	1440
GGATACCACC ACGGACGATT TCAGTCATGT AGGCACCGGT ATTGATTGAA ACGATGAAGA	1500
TAGCAGCCAG TGTACGGTCA AGGTTGATCC CGAAAGCTTG GGCAGTTCCA TAGTAGATAA	1560
CCATCGATTG AACAATCATT GCGGTACCAC GGAAAATTTC AATGTAGACA TTGAGAACCC	1620
AGCCGACTAG TTTTGTAGG CCGTAAATGA CTTTGTTC AGAGAGAGGA GCAGTACGGA	1680
AGACACCAAT GGCAAGTCCA ATAATGAGAC CTATCATGGT TCCGACGATA GAGATTAAAA	1740
GAGTGATACC AGCACCACGC AAGAGTTGTT GCCAGTTTC AGAAAGAATT TTAGCAACTT	1800
GGCTAAAGAA ACTACTGCTA GTCTCTTCAG TTGTTGTAGC TTCGGCAGGT TGTTCCTTGA	1860
TCATACGATC CATCAAGGCA ACTTGGTCAT CTTTGAAT GGTTCATG CTGGCATTGA	1920
TTTGGCTAAT ACGATTGTCA TTTTACGAA GCGCGATAGC GATAGCTGTA TCTTCTCCC	1980
CAGTTTTGAA ACCAGTTCT ACTTGAATCA TCTTGAATT AGAGTTCGCA GCTTCAGCAG	2040
TCAGTGCTTC TGGACGTTCA GAAACATAAG CATCAATGAC ACCAGCCTCA AGAGCTTGTC	2100
GCATTGAGC GAAGTCTCCC ATGGCTGTTT CTTTTTAGC ACCTGGGATT TGTGCAATCA	2160
AGTTATAAAG GTAGACCCCT TGTGAGAAG TGATTTTGC ACCGTAAAG TCATCCAAAG	2220
ATTTAGCACT TCGTAGGCA GAATCTTTT TGACAAGCAA AACTGGTTCG CTAGTATAGT	2280
AACTGCTCGA AAAGGCAATT TCTTGTTCG GTTCTGCAGT TGGACTCATA CCTGGGATAA	2340
TCATGTCAAT CTTACCAGAA GTAAGGGCAG GGACTAGACC TTCCCACTTG GTTTTAACAA	2400
CCAAAGGTTT TTTACCTAAG TCCTTAGCGA TTTCTTGGC GATTTGAACA TCGTATCCGT	2460
TGGCATACTG ATTGGTCCCA TCGATTTGA CAGCTCCGT GCTATCATCA TCCTGGGTCC	2520
AGTTAAAGGG AGCATATGCT GCTTCCATAC CGATGCGTAA ATATTCATCG GCTTGAGCAA	2580
CATTGACAAG TCCTAGCATC AGCAAGAGAC TTGTGAAAAT AGATAAGTAY ATGTGGCTCA	2640
TGATTTCTCC TATTCTGATC TATTAAAAA TAACTGTCTC CTATTTTATC GAAAAATGCG	2700
TAATTTTCA ACATAAGTAA GTCTTTACTT ACGAAAAAAT GCTATAATGA TAAGAAAGAT	2760
AAAAAGGGGG CTTAGTTGAT GAAAAAACT TTTTCTTAC TGGTGTAGG CTTGTTTTCG	2820
CTTCTCCAC TCTCTGTTT TGCCATTGAT TTCAAGATAA ACTCTTATCA AGGGGATTTG	2880

480

TATATTCATG CAGACAATAC GGCAGAGTTT AGACAGAAGA TAGTTTACCA GTTTGAGGAG	2940
GACTTTAAGG GCCAAATCGT GGGACTTGA CGTGCTGGTA AGATGCCTAG CGGGTTTGAC	3000
ATTGACCCTC ATCCAAAGAT TCAGGCCGCG AAAAACGGTG CAGAACTAGC AGATGTGACT	3060
AGCGAAGTAA CAGAAGAAGC GGATGGTTAT ACTGTGAGAG TCTATAATCC AGGTCAGGAG	3120
GGCGACATAG TTGAAGTTGA CCTCGTCTGG AACTTAAAAA ATTTACTTTT CCTTTATGAT	3180
GATATCGCTG AATTAAATTG GCAACCTCTG ACAGATAGTT CAGAGTCTAT TGAAGAAGTTT	3240
GAATTTCATG TAAGGGGAGA CAAGGGGGCT GAAAACTCT TTTTCCATAC AGGGAAACTT	3300
TTTAGAGAGG GAACGATTGA AAAGAGTAAC CTTGATTATA CTATCCGTTT AGACAATCTT	3360
CCGGCTAAGC GTGGAGTTGA GTTGCAATGCC TATTGGCCTC GGACCGATTT TGCTAGCGCT	3420
AGGGATCAGG GATTGAAAGG GAATCGTTTA GAAGAGTTTA ATAAGATAGA AGACTCGATT	3480
GTTAGAGAAA AAGATCAGAG TAAACAACTC GTTACTTGGG TCCTCCCTTC GATCCTTTCC	3540
ATCTCCTTGT TATTGAGTGT CTGCTTCTAT TTTATTTATA GAAGAAAGAC CACTCCTTCA	3600
CTCAAAATATG CCAAAAATCA TCGTCTCTAT GAACCACCAA TGGAAATTAGA GCCTATGGTT	3660
TTATCAGAAG CAGTCTACTC GACCTCCTTG GAGGAAGTGA GTCCCTTGGT CAAGGGAGCT	3720
GGAAAATTCA CCTTTGATCA ACTTATTCAA GCTACCTTGC TAGATGTGAT AGACCGTGGG	3780
AATGTCTCTA TCATTTTCTA AGGAGATGCA GTTGGTTTGA GGCTAGTAAA AGAAGATGGT	3840
TTGTCAAGCT TTGAGAAAGA CTGCCTAAAT CTAGCTTTTT CAGGTAAAAA AGAAGAACT	3900
CTTTCCAATT TGTTCGCGGA TTACAAGGTA TCTGATAGTC TTTATCGTAG AGCCAAAGTT	3960
TCTGATGAAA AACCGATTCA AGCAAGAGGG CTTCAACTCA AATCTTCTTT TGAAGAGGTA	4020
TTGAACCAGA TGCAAGAAGG AGTGAGAAAA CGAGTTTCCT TCTGGGGGCT CCCAGATTAT	4080
TATCGTCCTT TAACTGGTGG GGAAAAGGCC TTGCAAGTGG GTATGGGTGC CTTGACTATC	4140
CTGCCCCAT TATTCGGATT TGGTTTGTTC TTGTACAGTT TAGACGTTCA TGGCTATCTT	4200
TACCTCCCTT TGCCAATACT TGGTTTCTA GGGTTAGTTT TGTCTGTTTT CTATTATTGG	4260
AAGCTTCGAC TAGATAATCG TGATGGTGT CTAAATGAAG CGGCAGCTGA GGTCTACTAT	4320
CTCTGGACCA GTTTTGAAAA TATGTTGCGT GAGATTGCAC GATTGGATCA GGCTGAACTG	4380
GAAAGTATTG TGGTCTGGAA TCGCCTCTTG GTCTATGCCA CCTTATTGG CTATGCGGAC	4440
AAGGTTAGTC ATTTGATGAA GGTTCATCAG ATTCAAGTGG AAAATCCAGA TATCAATCTC	4500
TATGTAGCTT ATGCCGCGCA CAGTACGTTT TATCATTCAA CAGCACAAAT GAGCCATTAT	4560
GCTAGTGTCT CAAATACAGC AAGCACCTAC TCTGTATCTT CTGGAAGTGG AAGTTCTGGT	4620
GGTGGCTTCT CTGGAGCCGG AGGTGGCGGC AGTATCGGTG CCTTTTAAAG AGAGCTACCA	4680

481

TAGACTGAAA AAGTATGATA TAATGGAAGA TAGAAAAAAG ACAAACATA AGAAAAGTCA	4740
ATAGTTTTAT CTAAACTATT TCTTATTCA ATTTGATGAT TTGGCGATGA TTTTAGAGCA	4800
CGGCAAAAAG CCCTTGAAAA AGTCCATTTT TTCAAAGGTA ATCCTGTGTT AATTCAGAA	4860
ATTACATCAC TTTTGTTCG TCAAATGGCA GCTCTTTTTT AGGATATAAA ACAGGGTTCG	4920
GATAAGTTTT TTTGCAAGGT GGATGATGGC TACATTGTAA TGTTCCTT ATTCTAACTT	4980
AGTCTTAAGA TAGGCCTTAG AAGCAGGTGA AAAGCGAGGG CATGCTTTGG CAGCTTGAT	5040
GAGTGCCAC CGCAGATGAG GGAACCCCG TTTGACCATT CTCCAGCTA AATCAATCTG	5100
ACCTGACTGA TAAATAGAAG AATCCAGTCC AGCGAAAGCT TGTAAATGAG CAGGATTATC	5160
AAAGGCATGA ATATTTCGAA TCTCGGCTAA AATGACCGCC CTAAACGATC CCCAATCCCA	5220
GTAACCGTCG TGATGACCGA GTTGAACCTCA GCCATCGAGT CATTGATACA TGTTCGGCC	5280
TTGTCAATGA GCCTCTTGTA ATGCTTGATG ATTTCGAATT CACGAGCAGG AGATGTTGTT	5340
CCGATAGAAC GAGGTGCGAC TGAGAGGATA TCCTGAATTT TAGAAGCGGT CAATCGCTTA	5400
ATTTCTATCA GCTTATCAAA TCCTGCCTCA ATCCTTTTCT GAGGATTAGG GTAGCGTGTC	5460
AAGAGTTGGT AGGTATATTC TGAATGCTTT CCAACGATTT TATCCAATC AGGAAAGATG	5520
ATATCAAGAC AACGAGTGTA TTGTACTTTC CAATCAGACT GTTTTCTTG AGACGATGAA	5580
TATGTCTAGC CAGTATTTTT AGGTCTACTT GCCGATTATC GTGTTGAAAT TGTTACAGAT	5640
TGGGCTCAGA AAGAAGTTTA AGAGCGATGC CATGAGCGTC TTCTTATCC GTTTTAGTCT	5700
TGCGAAGTCA TAATGATTTG GCAAATTCCT TGATGAGCAA AGGATTGTAG GTGTAACTT	5760
TATATCCTTG TTCATGCAGG AAGTTCAGTA GATTAAAGGC ATAATGTCCA GTATCTCAA	5820
GAGCGATGAG ACAGTCTTGG TTGATCTGTC GAATAGACAG ATCTAAGAGT TCAAAACCAG	5880
CTTTATTATT TGAAAAAGTG AGTGGTTTAA GAACAGTTTT TCCTGGAACA TTCAAGGCTG	5940
TAACATCGTG TTTATTTTTA GCGATATCAA TGCCTACATA AAGCATGGGA GTACCTCCAG	6000
ATATAGTATT TCAAGTCTAC TTGGTTATCC ACGAATTTTT TGCCTTGTA CCTTAGACGA	6060
GATCAAACGT CTATGCGTTA TCAAATCAT TACCAATTGA AACAAAAGCT GTGGTTAGAG	6120
CCTTTCGGAA ATCGTCAAGC GATTGGAGGA AATGAACTAA TCCATAGTGG CTATTCCAA	6180
GTATACCACT TGGGCTTTGG CAGTAGCTAA CTGCGCTAAA TATAATATAG GGAGTAATCT	6240
ATGTATCTTA TTGAAATTTT AAAATCTATC TTCTTCGGAA TTGTTGAAGG AATTACGGAA	6300
TGGTTGCCGA TTTCCAGTAC AGGTCACTTG ATTTTAGCAG AGGAATTCAT CCAATACCAA	6360
AATCAAAATG AAGCCTTTAT GTCCATGTTT AATGTCTGTA TTCAGCTTGG TGCTATTTTA	6420

482

GCAGTTATGG	TGATTATTT	TAACAAGCTC	AATCCTTTTA	AACCGACCAA	GGACAAACAG	6480
GAAGTTCGTA	AGACTTGGAG	ACTATGGTTG	AAGGTCTTGA	TTGCTACTTT	ACCTTTACTT	6540
GGTGTCTTTA	AATTGATGA	TTGGTTTGAT	ACCCACTTCC	ATAACATGGT	TTCAGTTGCT	6600
CTCATGTTGA	TTATCTACGG	GGTTCCTTC	ATCTATTTGG	AAAAGCGCAA	TAAAGCGCGT	6660
GCTATCGAGC	CAAGTGTAAC	AGAGTTGGAC	AAGCTTCCTT	ATACGACCGC	TTTCTATATC	6720
GGACTCTTCC	AAGTTCTTGC	TCTTTTACCA	GGGACTAGCC	GTTCAGGTGC	AACGATTGTC	6780
GGTGGTTTGT	TAAATGGAAC	CAGTCGTTC	GTGTGACAG	AATTTACCTT	CTATCTTGGG	6840
ATTCTTGTTA	TGTTTGGAGC	TAGTGCCTTA	AAGATTTTCA	AATTTGTGAA	AGCCGGAGAA	6900
CTCTTGAGCT	TTGGGCAATT	GTTTTTGCTC	TTGGTCGCGA	TGGGAGTAGC	TTTTGCGGTC	6960
AGCATGGTGG	CTATTCGCTT	CTTGACCAGC	TATGTGAAAA	AACACGACTT	CACCCCTTTT	7020
GGTAAATACC	GTATCGTGCT	TGGTAGTGTT	TTGCTACTTT	ACAGTTTGT	CCGTTTATT	7080
GTATAAGAAA	AACCTTGAAG	GGGCAACTCT	TCAAGGTTTT	ATACTCTTCG	AAAATCTCTT	7140
CAAACCGCGT	CAGCTTTATC	TGCAACCTCA	AAACAGTGTT	TTGAGCAGCA	CTGCGGCTAG	7200
CCTCTAGTT	TGCTCTTTGA	TTTTTATTGA	GCTTTAAAA	CCAGTCATGG	TAATCCCCAA	7260
TAGGCGGACA	CCTCTTTCTT	TCTTGCTTAA	TTCTTCATAG	AGTTGCAGGG	CTATTTGGCT	7320
TATCTGACTA	GCATCTTGTG	TTTTTTGAGC	AAGACTTTTT	CGTTTGGTAA	GAGTTGAAAA	7380
GTCCTCGTAG	CGGATTTTCA	AAATGACAAT	TTTTCCAGCT	TTTTCTTGTT	GATGTAGATT	7440
GAGAGCGACT	TTTTCTGATA	GAAGAGTCAG	CTCTTTTTTG	ATATCTTCCT	CAGCAAGGAG	7500
AATCTTCCCG	TAGGTTTTCT	CCTTGCCGAT	TGATTTACGG	ATGCCAATGG	ATTTGACTGG	7560
AGAGTTGTGA	ATGCCACGAG	CCTTTCGATA	CAGATCATAG	CCTAGTCTAC	CAAAACGGTC	7620
TATTAGGGTT	ACCTCAGGAA	CTTCAAGTAA	ATCAGCACCA	GTAAAAACGC	CCATTTGATG	7680
AAGACGTTCT	ACTGTCTTTT	TTCTACTCC	ATGAAATTTG	GAAATATCCA	TTTGTTCGAG	7740
AAAATCCTCA	GCCTGTTTCA	GTAGAATCAC	TGTCAAACCA	TGTGGTTTTT	GATAATCACT	7800
CGCCATTTTA	GCTAAGAATT	TGTTGTAAGA	AACGCCTGCG	GAAGCAGTTA	GATGGAGTTC	7860
TTGCCAGATA	TCTTTTTGAA	TGAGGCGAGC	AATTTTGACC	GCTGACTTGA	TACCGAGTTT	7920
ATTTTCTGTC	ACATCCAAAT	AGGCTTCGTC	AATGCTCATG	GGTTCAATCA	AATCTGTATA	7980
GCGCTTAAAA	ATAGCTCGAA	TCTGGAGTCC	CACAGACTTG	TATTTCTCAT	AATTCCTGTA	8040
GATAAAGACA	GCCTGGGGAC	AACGTTTATA	AGCTTCCTTG	GAAGTCATGG	CAGAATGGAC	8100
ACCAAAAGCT	CTTGCCTCAT	AACACAGGT	AGAAACGACT	CCCCGTCCAC	CTGTTTGCCG	8160
AGGGTCGCTT	CCAATAATGA	CAGGTTTTCC	TCTGAGTTTA	GGATTATCCC	TGATTTCCAC	8220



483

TGCAGCAAAA AAGGCATCCA TGTCAATATG GATGATTTTT CTTGACAAAT CATTTAACAA	8280
AGGAAAAATC AACATGCCTA GCACCTTTTT ATACTCTTCG AAAATCTCTT CAAACCACGT	8340
CAGCTCTATm TGCAACCTCA AAACAGTGTT TTGAGCAATC TCGGGCTAGC TTCCTAGTTT	8400
GCTTTTCGAT TTCCATTGAG TGTTACTGCT TATTYTCTTT TATTATACCC TTTTCTGA	8460
AAAAAAGAAA AAAGGACTTT ATTTTTTCAA AAATATAATA CAGTTTGAAA TAAAATATAG	8520
ACTGTTTTAG AAAAGAAAGT GTAAAAATAG GGAATTTTCA CTTGTTGAAA TCGGTTACTA	8580
TATGGTATAC TTGTCTTATG AATGTAACAG ATGACTGTTA CTAGAAAAAA GAGGACATTA	8640
ATATGGTTGT TAAGACAGTT GTTGAAGCAC AAGATATTTT TGACAAAGCT TGGGAAGGCT	8700
TCAAAGGCGT AGATTGGAAG GAAAAAGCAA GTGTATCACC ATTTGTACAA GCTAACTACA	8760
CACCTTATGA TGGAGACGAA AGCTTCCTTG CAGGACCAAC AGAGCGTTCA CTTACATCA	8820
AGAAAATTGT AGAAGAAACT AAAGCACACT ACGAAGAAAC TCGTTTCCCA ATGGACACTC	8880
GTCCAACATC TATCGCTGAT ATCCCTGCTG GATTATCGA CAAAGAAAAT GAAGTTATCT	8940
TCGGTATCCA AAACGATGAA CTCTTCAAAT TGAACCTCAT GCCAAAAGGT GGTATCCGTA	9000
TGGCTGAAAC TACTTTGAAA GAAAATGGAT ACGAACCAGA CCCAGCTGTT CACGAAATCT	9060
TCACTAAATA TGTAACAACA GTTAACGACG GTATTTTCCG TGCCTACACT TCAAATATTC	9120
GTGCGCTCG TCACGCACAC ACTGTAAGTG GTCTTCCAGA TGCATACTCA CGCGGACGTA	9180
TCATCGGTGT TTACGCACGT CTTGCTCTTT ACGGTGCAGA CTACTTGATG CAAGAAAAAG	9240
TAAATGACTG GAATGCAATC AAAGAAATCG ATGAAGAAAC AATCCGCTCT CGTGAAGAAG	9300
TAAACCTTCA ATACCAAGCA TTGCAACAAG TTGTTGCGCT GGGTGACCTT TACGGGGTTC	9360
ATGTTGCGAA ACCAGCGATG AACGTGAAAG AAGCAATCCA ATGGGTAAAC ATTGCTTTCA	9420
TGGCTGTCTG CCGTGTGATT AACGGTGCTG CTACATCTCT AGGTCGTGTA CCAATCGTAT	9480
TGGACATCTT TGCAGAACGT GACCTTGCTC GTGGTACATT TACTGAATCA GAAATCCAAG	9540
AATTCGTTGA TGATTTGTTT ATGAAACTTC GTACAGTTAA ATTTGCTCGT ACAAAGCTT	9600
ATGACCAATT GTACTCAGGT GACCCAACCT TTATCACAAC TTCTATGGCT GGTATGGGTA	9660
ACGACGGTCG TCACCGTGT ACTAAGATGG ACTACCGTTT CTTGAACACT CTTGACAACA	9720
TCGGTAACTC ACCAGAACCA AACTTGACAG TTCTTTGGAC TGACAAATTG CCATACAACT	9780
TCCGTCGCTA CTGTATGCAC ATGAGCCACA AACACTCTTC TATCCAATAC GAAGGTGTAA	9840
CAACAATGGC TAAAGACGGA TATGGTGAAG TGAGCTGTAT CTCATGCTGT GTGTCTCCAC	9900
TTGATCCAGA AAATGAAGAA CAACGCCACA ACATCCAGTA CTTGCGTGCT CGTGTAACG	9960

484

TTCTTAAAGC CCTTCTTACT GGTTTGAATG GTGGTTACGA CGATGTTTAC AAAGACTACA	10020
AAGTATTIGA TATCGAACCA ATCCGTGACG AAGTTCTTGA ATTTGAATCA GTTAAAGCGA	10080
ACTTTGAAAA ATCTCTTGAC TGGTTGACTG ACACCTTACGT AGATGCCCTTG AACATCATCC	10140
ACTACATGAC TGATAGGTAC AACTACGAAG CTGTTCAAAT GGCCTTCTTG CCAACTAAAC	10200
AACGTGCCAA CATGGGATTC GGTATCTGTG GATTGCTAA CACTGTTGAT ACATTGTCAG	10260
CTATCAAATA CGCTACAGTT AAACCAATCC GTGACGAAGA TGGCTACATC TACGATTACG	10320
AAACAATCGG TGACTIONCCA CGCTGGGGTG AAGATGACCC ACCTTCAAAC GAATTGGCAG	10380
AATGGTTGAT CGAAGCTTAC ACAACTCGTC TACGTAGCCA CAACTATAC AAAGACGCAG	10440
AAGCTACAGT ATCACTTTTG ACAATCACAT CTAACGTTGC TTACTIONTAA CAACTGGTA	10500
ACTCACCAGT TCACAAAGGT GTATACCTCA ACGAAGATGG TTCTGTGAAC TTGTCTAAAC	10560
TTGAATTCTT CTCACCAGGT GCTAACCCAT CTAACAAAGC TAAAGGTGGT TGGTTGCAAA	10620
ACTTGAACTC ACTTTCTAGC CTGACTTTA GTTATGCAGC TGACGGTATC TCATTGACTA	10680
CACAAGTATC ACCTCGCGCT CTTGGTAAGA CTCGTGATGA ACAAGTTGAT AACTTGGTAA	10740
CAATTCTTGA TGGTTACTTC GAAAACGGTG GACAACACGT TAACTTGAAC GTTATGGACT	10800
TGAACGATGT TTACGAAAAA ATCATGTCAG GCGAAGACGT TATCGTACGT ATCTCTGGAT	10860
ACTGTGTAAA CACTAAATAC CTCACTCCAG AAAAAAAAC TGAATTGACA CAACGTGTCT	10920
TCCACGAAGT TCTTTCAATG GATGACGCCT TGGATGCATT GAGCTAATCA AGTTCTTGAA	10980
TAATAAAAAG GAACCTTCGG TCAAACGACT GAGGGTTTTG TGCTTGGGAT AGTATGAGCA	11040
ATTCCCTTCGG CGCAATATGC AATGTTTTTG GGCTCTTTGT CAACTGTAGT GGGTTGAAAA	11100
AAAGCTAAGC TTGAGAAAAG ACAAATTTTC TCCTTTCTTT TTTGATGTTT AGGGCGATAA	11160
AAATCCGTTT TTTGAAGTTT TCAAAGTTCC GAAAACCAAA GGCATTGCCG TTGATGTCTT	11220
TGATGAGTTT GTTAGTGGCC TCAAGTTTAG CGTTAGAATA AGGCAATTCA ATGGCGTTAG	11280
TGATGTAGTT TTTATAGCAA ATAAATGTGC TCAAAGTGGT TTTAAAGGTG CGGTGAGAT	11340
GAGGTAACGT GTCTTGAATT AAGCCCCAAA ACTGGTCAGT ATTCTTCTCT TGTAGATGAA	11400
ATAGGAGTAG TTGATACAGG TCATAGTAAT CTTTAAAGTT AGGTACTAGA GTAAAGATTT	11460
TCTTCAGACA CTCCCTAGGA GTTAAGGTCT CTCTGAAAGT TCTAGCATAG AAAGGCTTAA	11520
GAGAGAGTTT CCGACTATCT TTTAGGATAA ATTTCCACTA ATATTTAAGA GCTCTGTATT	11580
CCAGAGATTT ATCATCAAAT TGCTTCATGA TGTGATTTCT AGTCTGATTA AGAGCCCTGC	11640
TCATGTGTTG GACAATGTGG AAACGATCGA GAACAATTTT AGCATTGGGA AATAATTTCT	11700
TAAATGAGAGG GATATAACTT CCAGACATAT CAACAGTGAC GACTTTAACT TTTTTCTAG	11760

485

CTTCTTTTGA GTACTTGAAG AAATGATTTC GGATGGTTGT TTGACGTCTG TTATCAAGAA	11820
TGGTCATGAT TTTCTTAGTG TTGAAATCCT GAGCAATGAA AGCCAATTTT CCCTTCTGGT	11880
AGGAGAATTC ATCCCAGGAG AGGATTTTCAG GCAAAGTGGT GTAATCCTCT TGGAAATGAA	11940
ATTGCTTGAG CTTACGATAG ACGGTAGAGG TAGAGGTAGA GGTAGAGATG GCTAATTTAG	12000
AAGCGATATG TGTAAGAGCC TCTCTGTTGA GTAGGAGTTG GGCAATTTTC TGTCTCACCA	12060
TTTCCGAGAT TTGGCAATTT TTCTGAACGA GAGTTGTTTC AGCTACAGTG ACTTCCGAC	12120
AGGACTTGCA TTGAAATCGT CTCTTTTTCA AATGAATGAG GCTAGGGAAA CCACCAATCT	12180
CGATAAAGG GATTTTAGAA GGCTTTTGGG AGTCGTATTT GATTGTTTTT CCTTTACAGT	12240
GTTTACATTT AGGTGGGTGA TAATCAAGTG TAGCGAAGAC TTCGATATGG GTATCGTGCT	12300
GAATGGCTTT ATTTAAGGTG ATGTTTTTGT CTTTTATTCC GATGAGTAAT GTGGTATGAT	12360
TGATGTGTTT CATAAGATAC TTTCTAATGA GTTGTTTAGG CGCTTTTCAT TATAAGTCTT	12420
ATGGGACTTT TTTGATACTC AAAAAGCCCT ATAATCTCCA CAGTGGGATT TACCCACTAC	12480
AGAAATTATA GAGCCAGAAA AAACACTTTT GTTCACTAGC AGAAACTAGA GAGCAGAAGT	12540
GTTTTTCTGT TCAGATTTAC CCAAACTGG GAAATATGGG GATAAGAATA GAGATGGCTT	12600
AGGAAGCCCC TTTTGTGTG TAGACAGTAC GATGAACCTA TAACAAATAG TGAGCCTTTT	12660
TAGCAATCAT TGCGACCCGT TTGTCAAAAAG CCTCTTTTCG GATATCTACA ATTGTCTGAT	12720
AGATGAGACG CTGTTGGCTA ACATGCAAAT CTAAGGCAAT CGTCAAAAAG TGATGTTTCC	12780
CTTTGGGATA CTGCTTTTTA ACGTAAGGCA GGTATTCTTT CGTTGTAATA ATAATCAATG	12840
GCTCTGTCAA ATGCTCCTCT GAAGGAGGAG GACTAATTAG AATATTGTAT CCTGTAACAG	12900
AGGCAACTTT GTCAGTAAAA TTCCGTAAAA TAATGGACTT TATTAAGTTT ACATCTGCTT	12960
GATTATTTAA AATGATAAAA ATCGGGATAG CAGGTAGTGA GGAAAAGATG GTTTCTGTCA	13020
AGTAGAGTGA GAAAAGGTAC AGCCGATGCT GGTGATAAC TCCTTCAATC TTCTGCTCAG	13080
TCATCCACTC TTGAACAATT GCTTTGAAA TATGATACAG TGGCTTGTCT CTTTCAATCC	13140
CATAATGTTT GTAATAATTA TAATAGGGAA CTAGATTTTG TAAACCAAAC AAAACGTTT	13200
TTGTTAAGAA AGTCAGTGCT GTTAAAAAG AAAGAGAATT CGAAATGTCA TTTCCTAAGA	13260
TATCTTGAA CTGGATAGT AGATGCTTTC CTCTTGATG CTGAAGAATC AGTTGAATAG	13320
TATGAGTCTT TTTTCTTGA TTCCATTTGT CCTTGGAAAA CGAAGAATTA GCAGAACAAT	13380
AAACCAAAAA GATATAATCC AGTCTTCTCT GAGTAAAAGT CATGTTGGCA TGTGGCTCTA	13440
AGTAAGTTTG CCAATGTTCC ATCAAAATCG GATACATAAA GAGGTTTTTT AATTTTTCAA	13500

486

ACTCTTTGGA CTCAGGGAAC TCAAGTGGA	ATTCCCGACG TTTCCAAGTG AGTGCCACTA	13560
GTATGCTAAA ATGAACATAC TCGTCAGGTG	TGATTTCTAA CAGTTCATGA CTGAGTTGAG	13620
AATTAGACTG CACAATCATA TGTGTGACCC	AATCCATACT TCCATCATTC AAATCATAAA	13680
TCTCAATACC AAAATGAAAC TGGAGGAGTG	CAATTAAAAA ACGAATGCCA TATTCAGGAC	13740
CAACTACTTG ATTTTTCACA AGGTCCAAAC	CTACTGAACG TAGTAACAAG CCACACTTTT	13800
GTCGTACGCG GTAGCCTGTT GCGATGGAAA	TATACTCTTT TTGTGTAAAT TCGTTAAAGC	13860
TTTGATTACC TTGTAGTAGA AAGAAGCGGA	GTATTTTAA AATAGTTGAT TGGTTATAAA	13920
GCTGATGGAA GTAATAATTC GTTTGATGAG	AATGGTGTTC GATTAATTGA ACTTGTTGCG	13980
TATCTAAATT AAATGTCAAC TCTTCCTCGA	ATGTTTCTTG TAATTCCTGC AAAATGCTTA	14040
GGAGACTTTT AGATTGTAAT GAAGTTAAAG	TAGACAGTTC ATCTAGTTCA ATAGACCGAA	14100
TATCCAATAA TATATTTAAA ATGGTAATTT	TATCTGTAAT TCTTTTTTCA ATGTATTGT	14160
TTAGCATAGT TACCGAATCT TAGTTGCATA	TAGATAATTT TAATTATTAT AATACAAAAG	14220
AAACTAATTG TCTTGTCAAA AAGGTTGTGG	AATTTCCGAC TTTATTGATA AAACAGCATG	14280
TAATAAAAGG CATTTTAAAG ATAGTAATGA	GTATTGGTGG AGTTTTATGG CTTATTTTTT	14340
TTATTAGAAA ATATTTTTTT ATCAAATATT	GTCTTCTAT AAAAAAATAT GTGATAAAAA	14400
TATCTATTGT GATGGAAGTT GTTTTAATTT	ATACTAGGAT AGTTAATAGT AATACTATAC	14460
TATACTATAT TGTATACAAG TGTGTCATTG	CCAGGTTGAG AAGATAGCTA TAACGCACTT	14520
TTATACGCTT TTGCTACGTT TGTTAGTGAA	CGGATTAAC CAGTGAGATA AATTTTATCA	14580
GAACATAAGT AATCCGTTTC TTCGTGTATA	CAGATTGAAA GTACCTATGA ATCATAGAAG	14640
GATTAACCTG TTCTATGAAT AATGCTTAAC	AGGGAGACAC ACATGAAAAA AGTAAGAAAAG	14700
ATATTTCAGA AGGCAATTGC AGGACTGTGC	TGTATATCTC AGTTGACAGC TTTTCTTCG	14760
ATAGTTGCTT TAGCAGAAAC GCCTGAAACC	AGTCCAGCGA TAGGAAAAGT AGTGATTAAG	14820
GAGACAGGCG AAGGAGGAGC GCTTCTAGGA	GATGCCGTCT TTGAGTTGAA AAACAATACG	14880
GATGGCACAA CTGTTTCGCA AAGGACAGAG	GCGCAAACAG GAGAAGCGAT ATTTTCAAAC	14940
ATAAAACCTG GGACATACAC CTTGACAGAA	GCCCAACCTC CAGTTGGTTA TAAACCCTCT	15000
ACTAAACAAT GGACTGTTGA AGTTGAGAAG	AATGGTCGGA CGACTGTCCA AGGTGAACAG	15060
GTAGAAAATC GAGAAGAGGC TCTATCTGAC	CAGTATCCAC AAACAGGGAC TTATCCAGAT	15120
GTTCAAACAC CTTATCAGAT TATTAAGGTA	GATGGTTCGG AAAAAACGG ACAGCACAAG	15180
CGGTTGAATC CGAATCCATA TGAACGTGTG	ATTCCAGAAG GTACACTTTC AAAGAGAATT	15240
TATCAAGTGA ATAATTTGGA TGATAACCAA	TATGGAATCG AATTGACGGT TAGTGGGAAA	15300

487

ACAGTGTATG AACAAAAAGA TAAGTCTGTG CCGCTGGATG TCGTTATCTT GCTCGATAAC	15360
TCAAATAGTA TGAGTAACAT TCGAAACAAG AATGCTCGAC GTCCGGAAAG AGCTGGTGAG	15420
CGGACACGTT CTCTTATTGA TAAAATTACA TCTGATTACAG AAAATAGGGT AGCGCTTGTG	15480
ACTTATGCTT CCACTATCTT TGATGGGACC GAGTTTACAG TAGAAAAAGG GGATAGCAGAT	15540
AAAAACGGAA AGCGATTGAA TGATTCTCTT TTTTGGAAAT ATGATCAGAC GAGTTTACAA	15600
ACCAATACCA AAGATTATAG TTATTTAAAG CTGACTAATG ATAAGAATGA CATTGTAGAA	15660
TTAAAAATA AGGTACCTAC CGAGGCAGAA GACCATGATG GAAATAGATT GATGTACCAA	15720
TTCCGTGCCA CTTTACTCA GAAAGCTTTG ATGAAGGCAG ATGAGATTTT GACACAACAA	15780
CGGAGACAAA ATAGTCAAAA AGTCATTTTC CATATTACGG ATGGTGTCCC AACTATGTCTG	15840
TATCCGATTA ATTTAATCA TGCTACGTTT GCTCCATCAT ATCAAAATCA ACTAAATGCA	15900
TTTTTTAGTA AATCTCCTAA TAAAGATGGA ATACTATTAA GTGATTTTAT TACGCAAGCA	15960
ACTAGTGGAG AACATACAAT TGTACGCCGA GATGGGCAAA GTTACCAGAT GTTTACAGAT	16020
AAGACAGTTT ATGAAAAAGG TGCTCCTGCA GCTTTCCAG TTAACCTGA AAAATATTCT	16080
GAAATGAAGG CGGCTGGTTA TGCAGTTATA GGCGATCCAA TTAATGGTGG ATATATTTGG	16140
CTTAATTGGA GAGAGAGTAT TCTGGCTTAT CCGTTTAATT CTAATACTGC TAAATTTACC	16200
AATCATGGTG ACCCTACAAG ATGGTACTAT AACGGGAATA TTGCTCCTGA TGGGTATGAT	16260
GTCTTTACGG TAGGTATTGG TATTAACGGA GATCCTGGTA CGGATGAAGC AACGGCTACT	16320
ACTTTTATGC AAAGTATTTT TAGTAAACCT GAAACTATA CCAATGTTAC TGACACGACA	16380
AAAATATTGG AACAGTTGAA TCGTTATTTT CACACCATCG TAACTGAAAA GAAATCAATT	16440
GAGAATGGTA CGATTACAGA TCCGATGGGT GAGTTAATTG ATTTGCAATT GGGCACAGAT	16500
GGAAGATTGG ATCCAGCAGA TTACACTTTA ACTGCAAACG ATGGTAGTCG CTTGGAGAAT	16560
GGACAAGCTG TAGGTGGTCC ACAAATGAT GGTGGTTTGT TAAAAAATGC AAAAGTGCTC	16620
TATGATACGA CTGAGAAAAG GATTCGTGTA ACAGGTCTGT ACCTTGAAC GGATGAAAAA	16680
GTTACGTTGA CCTACAATGT TCGTTTGAAT GATGAGTTTG TAAGCAATAA ATTTTATGAT	16740
ACCAATGGTC GAACAACCTT ACATCCTAAG GAAGTAGAAC AGAACACAGT GCGCGACTTC	16800
CCGATTCCTA AGATTCGTGA TGTGCCGAAG TATCCAGAAA TCACAATTTT AAAAGAGAAA	16860
AAACTTGGTG ACATTGAGTT TATTAAGGTC AATAAAAAATG ATAAAAAACC ACTGAGAGGT	16920
GCGGTCTTTA GTCTTCAAAA ACAACATCCG GATTATCCAG ATATTTATGG AGCTATTGAT	16980
CAAAATGGCA CTTATCAAAA TGTGAGAACA GGTGAAGATG GTAAGTTGAC CTTTAAAAAT	17040

488

CTGTCAGATG	GGAAATATCG	ATTATTTGAA	AATTCCTGAAC	CAGCTGGTTA	TAAACCCGTT	17100
CAAAATAAGC	CTATCGTTGC	CTTCCAAATA	GTAAATGGAG	AAGTCAGAGA	TGTGACTTCA	17160
ATCGTTCAC	AAGATATACC	AGCGGGTTAC	GAGTTTACGA	ATGATAAGCA	CTATATTACC	17220
AATGAACCTA	TTCTCCAAA	GAGAGAATAT	CCTCGAACTG	GTGGTATCGG	AATGTTGCCA	17280
TTCTATCTGA	TAGGTTGCAT	GATGATGGGA	GGAGTTCTAT	TATACACACG	GAAACATCCG	17340
TAAAGTGTAG	AAATGATAAT	ATCTATGTTT	TGAACGATAC	TTTAAAGAAG	TAGCACTCAA	17400
GAAGAGATTT	AAGTTTACTT	GGTGAAACCT	GTTTTATTTCG	TAAGTAAACT	ATCATTGAAA	17460
GGGGAGATGT	TTTCGAAAAC	TTGCACAGAA	AAAGGATTAT	TATTGTCATG	TGTAATTCAT	17520
TACATTGCTC	ACAGTTGATT	TTAAGAGATA	TGAATAAGGA	GAAATCATGA	AATCAATCAA	17580
CAAAATTTTA	ACAATGCTTG	CTGCCTTATT	ACTGACAGCG	AGTAGCCTGT	TTTCAGCTGC	17640
AACAGTTTTT	GCGGCTGGGA	CGACAACAAC	ATCTGTTACC	GTTCATAAAC	TATTGGCAAC	17700
AGATGGGGAT	ATGGATAAAA	TTGCAAATGA	GTTAGAAAACA	GGTAACTATG	CTGGTAATAA	17760
AGTGGGTGTT	CTACCTGCAA	ATGCAAAAAGA	AATTGCCGGT	GTTATGTTCC	TTTGGACAAA	17820
TACTAATAAT	GAAATTATTG	ATGAAAATGG	CCAAACTCTA	GGAGTGAATA	TTGATCCACA	17880
AACATTTAAA	CTCTCAGGGG	CAATGCCGGC	AAC TGCAATG	AAAAAATTAA	CAGAAGCTGA	17940
AGGAGCTAAA	TTTAACACGG	CAAAATTTACC	AGCTGCTAAG	TATAAAATTT	ATGAAATTCA	18000
CAGTTTATCA	ACTTATGTCG	GTGAAGATGG	AGCAACCTTA	ACAGGTTCTA	AAGCAGTTCC	18060
AATTGAAATT	GAATTACCA?	TGAACGATGT	TGTGGATGCC	CATGTGTATC	CAAAAAATAC	18120
AGAAGCAAA	CCAAAAATTG	ATAAAGATT?	CAAAGGTAAA	GCAAATCCAG	ATACACCACG	18180
TGTAGATAAA	GATACACCTG	TGAACCACCA	AGTTGGAGAT	GTTGTAGAGT	ACGAAATTGT	18240
TACAAAAATT	CCAGCACTTG	CTAATTATGC	AACAGCAAAC	TGGAGCGATA	GAATGACTGA	18300
AGGTTTGCCA	TTCAACAAAG	GTACAGTGAA	AGTAACTGTT	GATGATGTTG	CACTTGAAGC	18360
AGGTGATTAT	GCTCTAACAG	AAGTAGCAAC	TGGTTTGTAT	TTGAAATTAA	CAGATGCTGG	18420
TTTAGCTAAA	GTGAATGACC	AAAACGCTGA	AAAAACTGTG	AAAATCACTT	ATTCGGCAAC	18480
ATTGAATGAC	AAAGCAATTG	TAGAAGTACC	AGAATCTAAT	GATGTAACAT	TTAACTATGG	18540
TAATAATCCA	GATCACGGGA	ATACTCCAAA	GCCGAATAAG	CCAAATGAAA	ACGGCGATTT	18600
GACATTGACC	AAGACATGGG	TTGATGCTAC	AGGTGCACCA	ATTCCGGCTG	GAGCTGAAGC	18660
AACGTTTCAT	TTGGTTAATG	CTCAGACTGG	TAAAGTTGTA	CAAACGTAA	CTTTGACAAC	18720
AGACAAAAAT	ACAGTTACTG	TTAACGGATT	GGATAAAAA	ACAGAATATA	AATTCGTTGA	18780
ACGTAGTATA	AAAGGGTATT	CAGCAGATTA	TCAAGAAATC	ACTACAGCTG	GAGAAATTGC	18840

489

TGTCAGAAC TGGAAAGACG AAAATCCAAA ACCACTTGAT CCAACAGAGC CAAAAGTTGT	18900
TACATATGGT AAAAAGTTTG TCAAAGTTAA TGATAAAGAT AATCGTTTAG CTGGGGCAGA	18960
ATTTGTAATT GCAAATGCTG ATAATGCTGG TCAATATTTA GCACGTAAAG CAGATAAAGT	19020
GAGTCAAGAA GAGAAGCAGT TGGTTGTTAC AACAAAGGAT GCTTTAGATA GAGCAGTTGC	19080
TGCTTATAAC GCTCTTACTG CACAACAACA AACTCAGCAA GAAAAAGAGA AAGTTGACAA	19140
AGCTCAAGCT GCTTATAATG CTGCTGTGAT TGCTGCCAAC AATGCATTTG AATGGGTGGC	19200
AGATAAGGAC AATGAAAATG TTGTGAAATT AGTTTCTGAT GCACAAGGTC GCTTTGAAAT	19260
TACAGGCCCTT CTTCAGGTA CATATTACTT AGAAGAAACA AAACAGCCTG CTGGTTATGC	19320
ATTACTAACT AGCCGTCAGA AATTTGAAGT CACTGCAACT TCTTATTCAG CGACTGGACA	19380
AGGCATTGAG TATACTGCTG GTTCAGGTAA AGATGACGCT ACAAAGTAG TCAACAAAAA	19440
AATCACTATC CCACAAACGG GTGGTATTGG TACAATTATC TTTGCTGTAG CGGGGGCTGC	19500
GATTATGGGT ATTGCAGTGT ACGCATATGT TAAAAACAAC AAAGATGAGG ATCAACTTGC	19560
TTAAGTAAGA GAGAAAGGAG CCATTGATGA CAATGCAGAA AATGCAGAAA ATGATTAGTC	19620
GTATCTTCTT TGTTATGGCT CTGTGTTTTT CTCTTGATG GGGTGACAT GCAGTCCAAG	19680
CGCAAGAAGA TCACACGTTG GTCTTGCAAT TGGAGAACTA TCAGGAGGTG GTTAGTCAAT	19740
TGCCATCTCG TGATGGTCAT CGGTGCAAG TATGGAAGTT GGATGATTCG TATTCCTATG	19800
ATGATCGGGT GCAAATTGTA AGAGACTTGC ATTCGTGGGA TGACAATAAA CTTTCTTCTT	19860
TCAAAAAGAC TTCGTTTGAG ATGACCTTCC TTGAGAATCA GATTGAAGTA TCTCATATTC	19920
CAAATGGTCT TTAATATGTT CGCTCTATTA TCCAGACGGA TCGCGTTTCT TATCCAGCTG	19980
AATTTCTTTT TGAAATGACA GATCAAACGG TAGAGCCTTT GGTCAATTGTA GCGAAAAAAA	20040
CAGATACAAT GACAACAAAG GTGAAGCTGA TAAAGGTGGA TCAAGACCAC AATCGCTTGG	20100
AGGCTGTCCG CTTTAAATTG GTATCAGTAG CAAGAGATGT TTCTGAAAAA GAGGTTCCCT	20160
TGATTGGAGA ATACCGTTAC AGTTCTTCTG GTCAAGTAGG GAGAACTCTC TATACTGATA	20220
AAAATGGAGA GATTTTGTG ACAAATCTTC CTCTTGGGAA CTATCGTTTC AAGGAGGTGG	20280
AGCCACTGGC AGGCTATGCT GTTACGACGC TGGATACGGA TGTCCAGCTG GTAGATCATC	20340
AGCTGGTGAC GATTACGGTT GTCAATCAGA AATTACCACG TGGCAATGTT GACTTTATGA	20400
AGGTGGATGG TCGGACCAAT ACCTCTCTTC AAGGGGCAAT GTTCAAAGTC ATCAAAGAAG	20460
AAAGCGGACA CTATACTCCT GTTCTTCAAA ATGGTAAGGA AGTAGTTGTA ACATCAGGGA	20520
AAGATGGTCG TTTCCGAGTG GAAGGTCTAG AGTATGGGAC ATACTATTTA TGGGAGCTCC	20580

490

AAGCTCCAAC TGGTTATGTT CAATTAACAT CGCCTGTTTC CTTTACAATC GGGAAAGATA	20640
CTCGTAAGGA ACTGGTAACA GTGGTTAAAA ATAACAAGCG ACCACGGATT GATGTGCCAG	20700
ATACAGGGGA AGAAACCCTT GTATATCTTG ATGCTTGTTG CCATTTTGTG GTTTGGTAGT	20760
GGTTATTGTC TTACGAAAAA ACCAAATAAC TGATATTCAA TGTACATCAT TATGAATAGG	20820
ATAGCAGGCT GAAGGGAAGA CCAGAGTACT CTGAGGTGAT GTTAATCAGG AATCATGGTG	20880
ATGTGGCATG AATCATCAAT AACGGATATG AGGCTGGGCA GATTGTGCCA GCCTCATTGT	20940
GGGTTATTGT TTGTAAACG ATAGGACTGG TCTGGTAATC ATTPTA	20986

(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 21040 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

CCCAGCAAAA AGCCATCCGA AGATGACTTT TTTGCTATTT AATTTCTGTA TAAGTTACTT	60
CCAAGCCACG CTTAACAGCT GGACGATTGG CAATTTTTC TGCCCATTTT ACTAGATTTT	120
GATAACTTGA GGCATCCAAG AATTTTGCAG AACCTTGGTA AAGATTTCTT TGAACAACT	180
GTCCATACCA AGACCAGATA GCAATATCTG CAATCGTATA GTCATTGCCT GCAATATAAG	240
GTTTCTGAGC CAATTCCTTA TCCAATAAAT CCAACTGGCG TTTCACCTCC ATCGTAAAC	300
GGTTAATAGG ATATCCAAT TTTTCAGGAG CATAATTGAA GAAATGTCCA AATCCCCCAC	360
CTAGAAAAGG TGCTGCACCT GCTTGCCAGA ATAGCCAATT CAAAACCTTCT ACCTTTTCCA	420
CAGGATTACT TGGTAAAAAG GCTCCAAATT TCTCAGCAAG GTAAAGAAGA ATATGAGCAG	480
ACTCAAAGAC TCTTACGTTT TCAGTACCTG ACTGGTCCAA TAAGGCTGGA ATCTTGGAA	540
TTGGATTGAG CTTACAAAAG TCTGATCCGA ATTGATCCCC ATCCATGATA GCAATCTTAT	600
ACAAGTCGTA AGCCGCTTCC TTAAAACCAG CTTCTAGTAA TTCTTCCAAT AAGATAGTAA	660
CCTTCACACC ATTTGGTGTT CCCAGTGAAT AAAGCTGAAA AGCTTGTTCT CCTTTTGGCA	720
AGTTTTGTTC GAAACGGGCA CCTGCTGTTG GTCTGTTTAG CCCCCTAAAA GCTCCTTGAT	780
TACTAGCTTC ATCCTGCCAT ACGGTCGGTA ATTGATATGC TGACATCCGA AACCTCCCTT	840
AAATCGCATT CTTGTCAAAA CCGAGTTTGC GTTGAATAAA CTTAACGATT TCGACGATGA	900
TAATCATTGA GAAGCTTCCA GCCATAACAA TTCCCCATTG TGACAAGTCT AGTTTGGTTA	960
CGTGAAGAT TCCTTCAAGC GGTTCACAA CGATTGTTGC CATGAGAAGG ATAAAGGATA	1020



491

CCAAGATGGA CCAGTTAAAG GTCTTAGACT TGAATGGGCC AACTGTCAAG ATGGATTGGT	1080
AGACAGACTT GACATTGTAG GCATGGAAGA GCTGAATCAA ACCAAGGGTT GCAAAGGCCA	1140
TCGTTAGGGC ATCTGCATGA ATAGCATGAT TGTCACCCAC ATGAACTGGG TAAGCAATCG	1200
CAAGGCCATA AACACTCATA ACAAGAGCTG CTTGGAGTAC ACCTTGATAA ATGATAGAAC	1260
TCAAAACACC ACCTGAGAAG AAGCTTGCCT TGGCTCCACG TGGTTTATGA TTCATGACAC	1320
CAGGTTCGGC AGGTTCAACA CCAAGAGCGA TAGCTGGGAA GGTATCCGTT ACCAAGTTGA	1380
TCCACAAAAG ATGAACCGGC TGTAAGACAT CCCAACCAAA CAAGGTTGAT AGGAAGATGG	1440
TTAATACTTC AGCAGTATTA GCAGAAAGTA GGTACTGAAT AGTCTTTTGA ATGTTTGAGA	1500
AGACCTTACG TCCCTCTTCC ACTGCGACGA TAATAGTCGC AAAGTTATCA TCTGCAAGAA	1560
TCATATCAGA AGCCCCCTTA GAAACCTCTG TACCAGTGAT TCCCATACCG ATACCGATAT	1620
CGGCTGTTTT CAGAGCTGGC GCGTCATTGA CACCGTCACC TGTCATGGCA ACGACTTTAC	1680
CTTGTTTTTG CCAAGCCTTG ACGATACGAA CCTTGTGTTT TGGAGACACA CGGGCATAAA	1740
CAGAGTATTG ACCAACGACT TTTTCAAAT CTTCATCTGA CAGTTCATTG AGTTCAGCAC	1800
CAGTTAAAAC GTGACCTTCT GTATCGTTTG CGTCAATGAT TCCCAAACGT TTGGCAATGG	1860
CTTCCGCTGT GTCTTGGTGG TCACCTGTAA TCATAATTGG ACGGATTCCC GCTTCCTTAG	1920
CCACACGAAC AGCCTCAGCG GCTTCAGGAC GTTCAGGGTC AATCATCCCA ATCAAACCAG	1980
TAAAAATTAA ATCATTTTCA AGCTCTTCAG AAGTGAGATT TTCTGGAATA CTATCGATAA	2040
TCTTATAAGC ACCTGCAAGG ACACGCAAGG CTTGATGAGC CATTTTCAGAA TTGTTTGTAC	2100
GAATGAGATT TGTAACCTTC TCATCAATCG GAGCAATATC CCCAGCCTTA TCACGAAGAA	2160
GACAACGTTT TAAGAGTTGG TCTGGCGCAC CCTTGACTGC TACAAGGAAA CGACCATCTG	2220
GCAATGGGTG AACTGTTGAC ATGAGCTTAC GGTGAGATC AAATGGCAAT TCAGCTACAC	2280
GAGGATATTT CTCTAAGAAA CCTTTGACAT CATAGCCCTT GTCCAAGGCA TATTGGATAA	2340
AGGCTGTTTC GGTGGGTCA CCAATCAAGT TACCTTCCAC ATCGATTTTC GTATCATTGG	2400
CCAAGACAAC TGAACGAAGT AGTGGCATT TCAAGACCTAG TTCAATATCA TCAGCTGAGT	2460
CATGTAGAAC CGCATCGTAG AAGACTTTTT CGACTGTCA TTTGTTTATA GTCAGCGTAC	2520
CAGTCTTATC AGAAGCGATG ATTTCAAGTT AACCAAGTGT TTCAACTGCT GGCAACTTAC	2580
GAACGATGGA ATGTCGTTTG GCCAAAACCT GAGTACCAAG AGAAAGAACG ATGGTAACGA	2640
TAGCAGGAAG TCCTTCTGGA ATGGCTGCAA CGGCAAGGCG AACAGAAGTC AACAACTCAC	2700
CAAGTGGATT TTTCCCTTGA ATGAAGACAC CCACTACAAA AGTAACAAGG GCAATGACCA	2760

492

AGATAGCATA	GGTCAAGACC	TTAGAAAGGT	TGTTCAAATT	TTGTTTGAGT	GGTGTATCAG	2820
TCTCATCCGC	ATCTTGAAGC	ATACCAGCAA	TATGACCAAC	TTCAGTGTAC	ATACCTGTAT	2880
TGACAACAAC	ACCCATCCCA	CGACCATAGG	TTACGTTTGA	GTTTTGGAAG	GCCATGTTGA	2940
CACGGTCACC	AATACCAGCA	TCTGTCGCAA	GCTCGACTGA	CAAGTCTTTT	TCGACTGGTA	3000
CAGATTCACC	TGTCAAGGCT	GCTTCTTCAA	TTTTAAGAGA	GTTGGCTTCT	ATCAAACGTA	3060
GGTCCGCTGG	TACCACGTCA	CCTGCTTCAA	GGGCAACGAT	ATCGCCTGGT	ACCAATTCTT	3120
TAGAGTCAAT	CTCTGCCATG	TGTCCATCAC	GAAGAACGCG	GGCAACTGGA	CTAGACATGG	3180
ATTTGAGGGC	TTCAATAGCT	TCTTCAGCTT	TTCTTCTTGT	GTAAACACCA	AAGGCAGCGT	3240
TGATGATAAC	CACAGCTAGG	ATGATAATGG	CATCTGCGAT	ATCTTCCCCA	CCAGAAGTCA	3300
CGACTGACAA	GATTGctGCC	GCAACTAGGA	TGATAATCAT	CAAATCCTTA	AATTGCTCGA	3360
TGAATTTGAC	CAAGATTGAT	CGTTTCTCGC	CTTCTTCGAG	TTCATTGTGC	CCAAATTCGG	3420
CAAGGCGCTT	TTCCGCGCTCA	CTTGATGACA	AACCTTGCTC	GGTCGCATCC	ACAGCCTGCA	3480
AGACCTCTTC	AGGGCTCTGA	GTATAAAACG	CTTGGCGTTT	TTGTTCTTTT	GACATGTGTC	3540
TCCTCCTTGA	CATTGTGTGC	AAAACAGACT	CTCTTTCTGT	CATAGCTTTT	CACGACAAAC	3600
AAAAAGAAAC	CTGTTAATCA	TAACAAGTCT	CGCTGTTTAA	GATAGGGCCG	GAAAGCATAc	3660
TTTTCAGCAT	AAAATTCGGA	ATGACGACAC	TATCACAGGT	TTCTGCCAGC	TACTCCCTTG	3720
AGTAGTACCA	TTATACCAA	TTTTGGGGAG	TTTTCAAAGA	GTAAAAACTG	CCTTATTTGA	3780
ATTTTTCCTT	GAAAACCAGT	ATAATGGTAG	AATGCTATGT	GACTIONAAAG	GAAGTTGAAT	3840
GAAGCAATCT	ATCTCAAAATC	TCAAGTTAGC	TGAGCGTGGA	GCCATTATCA	GTATTTTCGAC	3900
CTATTTGATC	TTGTCTGCAG	CCAAATTAGC	AGCTGGTCAT	CTCCctCATT	CATCCAGTTT	3960
GGTGGCCGAT	GGTTTTAATA	ACGTATCGGA	CATCATTGGA	AATGTGGCCC	TCTTAATCGG	4020
GATTCGGATG	GCGCGCCACC	TGCAGACCGT	GACCACCGTT	TTGGTCATTG	GAAGATTGAA	4080
GATTTGGCAA	GCTTGATCAC	TTCTATCATC	ATGTTCTATG	TCGGTTTCGA	TGTTCTAAGA	4140
GATACCATTC	AAAAGATTCT	CAGTCGGGAA	GAAACGGTCA	TTGATCCTCT	TGGTGCAACT	4200
CTAGGAATCA	TTTCTGCAGC	GATTATGTTT	GTGGTCTATC	TCTACAAATAC	TCGCCTCAGT	4260
AAGAAATCCA	ACTCCAATGC	GCTGAAGGCA	GCTGCTAAGG	ACAATCTTTT	TGACGCTGTT	4320
ACCTCACTTG	GAACCGCCAT	TGCCATCCTA	GCTAGTAGTT	TCAATTATCC	GATTGTGGAT	4380
AAACTGGTTG	CTATCATCAT	CACTTTCTTT	ATCTTGAAGA	CTGCCATATGA	TATCTTCATC	4440
GAGTCTTCCT	TTAGTCTTTC	AGATGGCTTT	GACGACCGCC	TGCTCGAGGA	CTACCAAAAG	4500
GCTATCATGG	AAATTCCCAA	AATCAGCAAG	GTCAAATCCG	AAAGAGGTCC	CACCTACGGT	4560

493

AGCAACATCT ACCTGGATAT TACACTAGAG ATGAATCCTG ACTTGTCTGT TTTTGAAAGC	4620
CATGAAATCG CGGATCAGGT CGAGTCTATG CTGGAGGAGC GTTTTGGCGT CTTTGATAACC	4680
GATGTCCATA TCGAACCAGC ACCTATCCCT GAGGATGAAA TTTTAGACAA TGTCTATAAA	4740
AAATTGCTTA TGGGTGAACA ATTGATTGAC CAAGGAAACC AACTAGAAGA ACTCTTGACT	4800
GATGATTTTG TCTATATTCG CCAAGATGGA GAGCAGATGG ATAAAGAGGC TTATAAGACC	4860
AAAAAAGAGT TAAATTCCTG TATCAAGGAC ATTCAAATTA CTTCCATCAG TCAAAAAACC	4920
AAACTCATCT GCTATGAGTT AGATGGTATC ATCCATACCA GTATCTGGCG TCGCCACGAA	4980
ACCTGGCAAA ATATCTTTCA TCAAGAAACC AAAAAAGAT AGAGAAATCC TTTCATGAGA	5040
CGGGATTTTT CTATCTTTT ATACTCAATA AAAATCAAAG TGCAAATTAG GAAGCCGGTC	5100
ACAGGCTGTA CTTGAGTCGG CAATGTGAAG CCGACATAGT TTGCACTTTG ATTTTCGAAT	5160
AGTCTTAACT ATCAAATTC CTGAGATACT CATAGCCTTC GTATTTTTC AAGAGTGCTT	5220
CATTTTCTC ATCCAATTCT TTTTGGAGAG TAGCCAGCTT ACCAAAGTCA GAGCCGTTAG	5280
CCTGCATTC CTCTTCAATA GCAGCGATAC GTTTTCCAA GGTTCATA TCACCTTCAA	5340
TACTTGCCCA CTCCTGCTTT TCTTGGTAGG TCATGCGTTT CTTGTCTTCT CGAACCTTGA	5400
CCACTTTTTT CTTTTCGGCC TTTTGCACCT GATTGGCCAT ATCTGTTTCA AAAGCTTTTT	5460
CATCAAGATA GTCCGTGTAA TGACCAAAGA AAGGACGAAT CTTGCCATCC TCAAAAGCGA	5520
GAATCTTGGT CGCTACCTTA TCCAAGAAAT AGCGGTCGTG ACTGACTGTT AAAACGGGAC	5580
CTGCAAAACC TTGCAAGAAA TTCTCTAAGA CTGTCAAAGT TGCAATATCT AGGTCATTGG	5640
TTGGCTCGTC TAAAAGAAGA ACATTGGTT TTTCCAAAAG CAGTTTGAGG AGATAAAGAC	5700
GTTTTTCTC ACCCCCTGAC AATTTCTCAA TCAAAGTCCC ATGCGTGGAA GGTGGGAAGA	5760
GGAATTGCTC CAGCAACTCA GCGATGGAAG TCGTAGAACC ACCACTGGTC TTGACCTCCT	5820
CTGCCACTTC CTGCAGGTAA TTGATCACAC GCTTGCTTTC ATCCAAACCC TCAATTTGTT	5880
GAGAGAAATA GGCGATGCGA ACAGTTTCCC CAATCACAAC TTGTCCTGCT GTCGGCTCAA	5940
GACTTCCTGC AATCAGGTTA AGTAGGGTTG ATTTTCCAAC ACCATTGTCC CCAACAATTC	6000
CAATACGGTC TTTAGCCTGA ACTAAGAGAT TAAAATTTTG CAAAATGGGC TTATTTTCAT	6060
AGGCAAAGGA AACATCCTGA AACTCGATGA CTTTCTTCCC AATCCGACTG GTTTCAAAGT	6120
TCATAGTCAA GTCTGTCTCA GCACTACTGC CTGAAACTTC CTTTTTCAGA TCATGGAAAC	6180
GATTGATACG AGCTTGTTGC TTGGTCGCAC GCGCCTGCGG TTGTCTGCGC ATCCAGGCCA	6240
ATTCTTGTTT GTAGAGTTGT TCTTTTTTGT GAAGAAGAGC CGCGTCGCGC TCATCCTGTT	6300

494

CCGCCTTTAG GCGAACATAG TCCTGGTAAT TTCCCTGGTA CTCGGTCAAG CCTGCACGAT	6360
CCAACTCGAA AATCCGTGTT GACAAAGCGT CTAAGAAATA ACGATCGTGA GTGATAAAAA	6420
GGACGGTCTT CTTAGAATTT TTCAAAAAGA GGGTCAGCCA CTCAATAATC GCAATATCCA	6480
GATGGTTGGT CGGCTCATCC AAAAGCAAGA GGTGCTGGTT GCCAAGTAAG ACTTGTGCCA	6540
ACTGTACCCG TCTTCTCAGA CCACCTGACA ATTCCCCAAC AGGAGTAGAT AAGTCTTGAA	6600
TGCCCCAATTT GCTAAGAACG GTCTTGACCT GACTTTCGAT TTCCCAAGCT TGGAGAGACT	6660
CCATCTCTGC CATGACACGT TCCAAACGCG CCTGCTTGTC CTCACTATAG TCGAGCATAA	6720
TCAATTCATA CTCACGAATG AGCTGGATTT CCTTGAGTTC ACTAGATAGA ACCGTATCCA	6780
AAACTGTCTT TCTATCATCA AAATCAGGAT CCTGAGTCAA GTAACCAATC TGGTAATCAT	6840
TTTTAGCTGA AAAAGGACTG ACATCCCCAT CAAATCCAGA AACACCAGAA AGGACGTCCA	6900
AAAGGGTGGT CTGCCAGTC CCATTGACAC CGATTAAACC AATTCTGTCT AAGTCATGGA	6960
TAATAAAGGA AATATCCCTA AAAACGGTCT TGTCACCAAC GGATTTACTT AGTTTTTCAA	7020
CGATAAAATC ACTCATTITT TCTCCCTCAG GTAAGCATGG ATGGCTTCAC GATTATTCTC	7080
CAATTCTCCA TCGACAATGG CAACTCAAT CTCTGTTAAA ATCTCTCCCA AGTCTGGGCC	7140
TGGCTGATAG CCATATTCTT TGATCAAAAT ACCGCCATTA ATCTGAATCT CTTTCTTGTC	7200
ATGGATAGTC AAGCTTTGGT ATTTTTCTGT GATGGCTTGT GGGTTGACTT CTTTCTTG	7260
AGCTTGACGA AGATTTTCAG CCTGTAAAAG CAAATCTATG TCAAAGCGAT AACAACTCTG	7320
CTTGCTCAAT TCTCCATTTT CACGCAGAGC CAAAATAATC AGCAAATCCT GAACTTGCTT	7380
GGCAAACCTG CGTGAGGTCT TCCAAGATTT CAAAAATGAC TGCGCATTTT CAATCTCCAA	7440
AGCCCATAGT AAAGCCGCCC AGGCTTGTTT AGAGGATTCA AAAGTAAAAT CAGTCTCCAA	7500
ATCAAACAGT CTGTTGAGCT TGTCCTGGCT AGATGCCATA TCAGGGAGAT AGTCATAAGC	7560
TTGACTCTCA ATCATGGAAG CCAAGCCCTT TCTCCAAAAT GGAGCCAGCA AGAGTTTATC	7620
AAACTCGACG AAGGTACGCT CTACAGAAAT TTTCTCCAAA AGCGGCGTCA AGGTCTTCAT	7680
AGCTTTAAAT GTTCTGGCT CAAGTGCAAA ACCAAGACTA GCCTGAAAAC GGAAACCACG	7740
CATAATCCGT AAAGCATCTT CGTTGAAACG CTCACTAGCC ACTCCAACGT CTCGCAAGAC	7800
TTGCTTTTCC AAATCTCTA AACCATGGAA CAAGTCAACG ATTTCTCTCG TCTCATCCAA	7860
GGCAAAGGCG TTGACTGTGA AATCACGGCG TTTGAGGTCT TCTTCTAGCG ATCGTACAAA	7920
GGAAACCGCA CTGGGTCTGC GATAGTCCAC ATAGACATCC TCTGTCCGAA AGGTTGTTAC	7980
CTCATACTCC TCATCCCCAT CTAAGACCAA GACGGTTCCA TGCTCGATTC CGATATCGGC	8040
TGTTCCGGGA AAAATCTGCT TGGTCTCTTC TGGATAAGAA GACGTCGCAA TATCCACATC	8100

495

GTGGATAGGG CTATGGAGAA GGGCATCTCG AACAGAGCCC CCAACAAAAT AAGCCTCAAA	8160
GCCTGCTTCT TTAATTTTTT CTAATACTGG TAAAGCCTTC TGAAATTCAG AAGGCATTTG	9220
CGTTAATCTC ATAATAAGTG TTCTAATCCA TAGACAAGCT CATGACGCTT GACAACTTCT	8280
TTAATTCCCA AATTGACTCC TGTCAATGAAG GAGATGCGAT CATAGGAGTC ATGACGGAGG	8340
GTCAACCCCTT CTCCCTGATT GCCAAAGATG ACTTCCTGAT GAGCTACCAA GCCTGGCAAA	8400
CGAACTGAGT GGATGCGCAT ACCATCAAAG TCAGCACCAC GAGCACCAGC AATCAGCTCT	8460
TCCTCATCTG CTGCACCTTG CTGAATTGAC TCTCGAACCT CTGCCATCAA CTCAGCTGTT	8520
TTAATGGCTG TTCCACTCGG AGCATCCTTT TTCTTGTCAT GATGGAGCTC AATAATCTCC	8580
ACATTTGGGA AATATTTGGC AGCCTGCGTC GCAAATTGCA TGAGTAAGAC AGCACCCAAG	8640
GCAAAGTTAG GGGCAATCAG GCCACCCAAG TCTTGGGCAC GAGAAAATTC TTTTAGCTCT	8700
GCAATTTCTT CACTCGTGAA ACCAGTCGTT CCAACTACTG GAGCAAAGCC ATTTTCAAGA	8760
GCAAAACGTG TATTTTCGTA GGCAACAGCT GGAGTAGTAA AATCTACCCA GACATCCGCT	8820
TCAAAACCAG CTAAATCAGC CTTATCCTTG AAAACAGGAA TACCCTGCCA TTCTGACTCA	8880
GACTCAAAAG GATCCAAAAC TGCCACCAAG TCCAAGTCTG GATCAGTCAA TACCATCTGA	8940
CAAGCAGCCT GGCCCATCTT TCCCTTAAAA CCGGCAATAA TTA CTGGAAT ACTCATCTCT	9000
ACTCCTGTCT AAGATACAAA GTCCGTAAGA ACACAAAGTG AAAATAGGAA TTCCAATCAA	9060
GAAGTGTCTA CTTCTTGAA GAATATCTT TTTACACAG GGTTCAGGC CTGTTCAATT	9120
ATCAAGATAC AAAGGACCTT AGCTGCCTCT GAAAAATAGG GAATGGCACT GACTTTCCAC	9180
GAAAGGCAAG ACAGGCATCT TTTTCAAGA GGCAGGTACT CCGTGTTCAA TTTCTAAGAT	9240
ACAAGGCATC TTAAGTAGCC TAGAAGCGCC AACTAAATCA CTGGAATATA ACCCAGAGCA	9300
ATACTTCCTG CTCCTAGGTG CGTTCCAATG AACTACCAA ATGTAGCAAG TGAACATCC	9360
GAACCCAAGC CAAAATCAAG CAAGTGCTGA CGCAATTCCT CAGCCTTTTC AGGAGCATTC	9420
CCATGAATGA CAATGACCCG GTATTGACCT GAAGCCGTG TTTCTTGAT AATTTCAATT	9480
AAGCGCTTGG TGGCCTTCTT TTCAGTACGA ACTTTTTCGT AAACCTCAAT CACACCTTGA	9540
TCGTTAAAAA AAAGGATTGG CTTAATGCTA AGCAAATTGC CCAAAATGGC AGCCCCATTT	9600
GAAAGGCGTC CACCTTTTAC CAAATGATCC AAGTCATCTA CCATGATAAA GGCTGACGTA	9660
CGGCTGATTT GAATGGCTAG CTTATCCTGA ATGCTGGCAA AATCATCGCC CTGATCACGC	9720
CAATTAAAGA CGCTTTCAAC CATGATGCCT AGGGGAGCAC TTGTAATCAA AGTGTCTGGG	9780
AAAGCAATGG TTAAGCCCTC ATAGTCATCG ACCATATACT GGATATTTTG GTAAAAACCT	9840

496

GAAATTCAG AAGATAGGAA AAGCCCCAAG GCATGTGTAT AGCCTTGTTT TTTGAGCGAA	9900
GTTAAGATCT CATCTAACTT GGCAATACTT GGTGACTCG TCTTAGGCAA TTCAGAAGCC	9960
TGAGCCATTT TTTGGTAAAA TTCCTCAGCA GACAGATTGA TGCCTTCGAC ATATTCCTCA	10020
CCATCAATAT TGACAGGAAT ATCCAAGACA AACAAGTCTT CTCTTTGCAA GATCTCTGCA	10080
CTGAGATAAG CAGAGGAATC TGTGAAAACA GCTAATTTCA TATTAGAACT CCAAATTAAT	10140
TCCTGGTAAG TCTAATGCAA TTTCAGTCAC TTCGTAAGTC AAACGATTGA GCATGTTCAA	10200
ACATGGACGA GCCAAGGTTT CCACCTCTTC TTGGTTCAAT TCACTTGGTT CATTGACAAT	10260
ACGGCCATCG ATATGGTTTA CTTGTGAGAT TGTCCACTA ATGACAACT TATCAAATAC	10320
AATCATAAAG CTCAAGATGA CAATCAAGGA AGTCACCTGA TTTTCTTGGT CATGTTGGAG	10380
CAATTGGAAT TTCACATCCA CCTTGGTTTC AGGAGCTCCA TTTTCATTTT CCCATTCAAA	10440
ATTACGGCGA TCAAAATGAT ACTGACTAAC AAATCTTTGT TCACGTTTAA GATTCATGTC	10500
TTTCTCCATC GGCTACAATA TTATAAGCTA TTGTACCATA ATTTTTTATT TTCATCTAGT	10560
TTTCTAGGAT TTAGTCAATC CCAATTTAG CACGAACTAC ATCTGTGATG GTATCAACAT	10620
AGTAGTTTAC TTCTTCTGTT GTAGGCGCTT CTGCCATAAC ACGCAAGAGG GGTCTGTTC	10680
CACTTGGACG AACAAGGATA CGGCCGTTCC CCGCCATTTT TTCTTCCATC TTCTCGATGA	10740
TGGCCTTGAT AGCTGGCACT TCCATGGCCT TTTCTTTCAT GACGTTTTTC ACTCGGATAT	10800
TAACTAATTT TTGTGGATAA ATCGTTACTT CTGCCGCCAA CTCTGATAAG CTCTTACCAG	10860
TTTCTTTCAT GATTTTAGTC AATTGAACG CTGATAATTG ACCATCACCT GTGGTATTGT	10920
AATCCATCAA GATAACGTGA CCAGACTGTT CACCACCAAG GTTGTAGCCT GATTTTCTCA	10980
TTTCTTCAAC AACGTAGCGG TCACCAACTG CAGTAACTGC CTTGTTAATA CCTTCGGAT	11040
TCAAGGCCTT GTGGAAACCA AGGTTAGACA TAACAGTTGT CACAATTGTA TTTTGAGCCA	11100
ATTGTCTTTT TTCAGAAAGG TATTTTCCGA TGATGTACAT AATCTTGTCA CCATCAACGA	11160
TGTCACCATC CTCATCAACA GCAATCAAGC GGTCACTGTC TCCATCAAAG GCCAAACCAA	11220
TAGCTGACCC ACTTTCTTTG ACCACTTCTT GAAGGGCTTC TGGATGTGTT GAACCAACAT	11280
TAAGGTTGAT GTTAAGACCG TCTGGTGTTC CCCCATAAC CGTCAATTGG GCACCAAGGT	11340
CTGCAAAGAT TTGACGGGCA CTGGTAGAAG CTGCTCCATT AGCTGTATCC AAGGCAACCT	11400
TCATTCATC AAGAGGAGTT CCAGTTGAAA CAAGGTATCC TTCATACTTA CGCAAGCTTC	11460
TGGATAATCT ACCAAAATTC CTAAGCCTTC TGCACTTGA CGAGGAAGAG TGTCTTCTTC	11520
AGCATCTAGC AAGGCTTCAA TTTCTGCTTC TTTTCATCA TCTAGTTTGA AGCCATCACC	11580
GCCAAAGAAC TTGATTCCGT TATCAAGGGC TGGGTTGTGG CTAGCAGAAA TCATGACACC	11640

497

GGCACTTGCT CCTTCAGTTT CAACCAAGTA AGCTACTGCT GGTGTTGCAA GGACACCAAG	11700
TTTGTATACG TGAATCCCTA CTGAAAgGAG ACCTGCCACC AAGGCCGATT CCAACATTTC	11760
CCCTGAAATA CGTGTGTCAC GTCCTACAAA GACTTTCGGC GCTTCCGTTT CATGTTGACT	11820
AAGAACATAG CCTCCAAAAC GTCCTAGTTT AAAGGCTAAT TCTGGTGTTA GTTCTAGGTT	11880
AGCTTCTCCA CGGACTCCAT CAGTCCCAAA ATATTACCC ATTGTTATAA AATCCTTTTC	11940
TATTTTAAAT TCGTTTTGA ACTAGTTGCT TTCGTTGACG AAGATGTCTC CGATGAAC TG	12000
CTTGTA CTG AATTGATGT GCTTGA ACTT GGTGCTACTG GTTTGTAGT CACCTTCATT	12060
ATTGTATCAA ACGGAGTGAT AACTGCCGGT AAGACAACAC CATTGCCGTC GATTGCCTGC	12120
AAAGGTACTG AACCACTGTA ATTACCTGTT ATACGTTGCG TAGTTGGCAA AACAGCGATA	12180
ATCTTATCAA TTCTATCCAA TGTCTCTGG TCACTCGTAA TAGACACTTC TTTATCTGAC	12240
ACCATGACAT TTTCAATTG TACCCGACTA TCAATTTGAC TAGGGTCAAT CTCTGGTACA	12300
ATCTTTACCT TATCCTCTG AGCCTTCTTA CCAATCTTGA CTGTAATTTT TTGCGGAGTC	12360
GCCACAGCGG TCAGCCCAT TGGTAAATCT TCAATGCTCA AAGGAACTTC AATCGTTCCA	12420
ACACCGGCAT CTGTTAGGTC AGCAGTAACC TTGAATTTAC GTGTACTTTC TTGCATTTCA	12480
CTAGCTAGCG ATAGGCGATT TGCACCAGTC AAGACCAGT ATACTTCTGA AGCAAAACCG	12540
CTAATAAAAT ACTTATCACT ATTATAGCGT ATGTCAATAG GGACATTTGT TACTGTATTA	12600
GTATAGGTTT CCGTTTTTAC GTGCCAGCA CTGGTACTGT TTTGAAAATT CGTCGCCGTA	12660
GCATAGACAA ATAAGACACA AGCAAAAAG AGTGAGGATA TGATATATAA ACTATTTTTT	12720
TTTCATGTTTC CATCCTCCTA GCAATCGTTC TTTAAACTA AGACCCACTT CCTCTTTTGG	12780
AAGTAAGATT TCACGTAATT CTGTTTCAAA TTCATCAAGT GTTAGGTTGT GCTTAAACCT	12840
TCCATTATAG GTTATCGAAA TTCCTCCCGT TTCCTCTGAT ACGACAAAAG TCAAGGCATC	12900
TGAGACTTCT GATAAACCGA TAGCCGCCCG GTGCTGGTC CCAAATTCCT TGGAAATCCC	12960
TGTGTTTTTT GTCAAGGGCA GATAGGCAGA CGTCACAGCG ATACGTTCTT CTTTGATAAT	13020
CACCGCACCA TCATGTAGGG GAGTGTGGG AATAAAAATG TTAATGAGAA GTTCTGCAGA	13080
AATCTTAGCA TCCAAGGGAA TTCCTGTCGA AATATACTCC TGCAAGGTAC GTACACGCTG	13140
AATAGCAACC AAGGCCCGA TTTTACGAGG ACTCATGTAT TCAACAGACT TAACAAAGGC	13200
ACGAATCATC TGTTCTCAG CACTAATAGG GGCATTGGAA AAGAAATCTG TCGCTCTTCC	13260
CAAACGTTCC AAACCAGTCC GAATCTCTGG AGAGAAGATA ACAACCGCCG CAATAACCCC	13320
ATAAGTAATA ATTTGATTGA TTAACCAAGA AATCGTAGTC AAACCAATCA TATTTGCAAG	13380

498

GATTTGAGCT	AAAATAAACA	CCAAAACCTCC	ACGTACCAAA	ATCATAATCT	TGGTTCCTGC	13440
AATAGCTTTT	GTAAAATGGT	ATAAAATATA	AGCAACAATC	AAAATATCAA	TCAGATTGAT	13500
AGCTATCGTC	CATGGACTTG	CAAACAAACT	GGTCCAATAT	TGCAGATTGG	ATAATTGTTG	13560
AAAATTCATC	CCTGATATCC	TCCCTATCAA	AACACTTTTC	TCCTATTATA	CCATTTTCTG	13620
GCATTTT	CCCTATCCTA	GTCCATTTTA	CATTGAACAA	AAATATGATA	AAATAAACTG	13680
ACTAAAAAAA	ACAAAGGAGA	AACTATGTCT	CAACTCTATG	ATATTACCAT	TGTGGGTGGT	13740
GGTCCTGTGC	GGCTTTT	AGCCTTTTAT	GCCCACCTAC	GCCAAGCCAA	GGTTCAAATC	13800
ATCGACTCTC	TTCCCCAGCT	AGGTGGACAA	CCTGCTATTC	TCTACCCCTGA	AAAGGAAATC	13860
CTAGACGTAC	CAGGCTTCCC	AAACCTGACT	GGAGAAGAGT	TGACTAACCG	CTTGATTGAA	13920
CAGCTAAATG	GATTTGATAC	CCCTATTTCAT	CTCAATGAAA	CGGTTCTTGA	GATTGACAAA	13980
CAAGAAGAAT	TTGCCATCAC	AACTTCTAAA	GGAAGTCACC	TGACTAAAAC	AGTTATCATC	14040
GCTATGGGTG	GCGGTGCCTT	CAAACCACGT	CCGCTGGAAC	TTGAAGGGGT	TGAGGGCTAT	14100
GAAAAATATCC	ACTACCACGT	TTCTAACATT	CAGCAATACG	CTGGTAAGAA	AGTGACGATT	14160
CTTGGTGGGG	GAGACTCGGC	TGTGGATTGG	GCTTTGGCTT	TTGAAAAAAT	CGCACCAACT	14220
ACCCTTGTTC	ACCGCAGAGA	TAATTTCCGT	GCCTTGGAAC	ACAGTGTTCA	AGCCTTGCAA	14280
GAATCATCTG	TAACCATCAA	GACACCATTC	GCCCCTAGCC	AACTCCTTGG	AAATGGAAAA	14340
ACACTTGATA	AACTTGAAAT	CACAAAAGTC	AAATCTGATG	AAACTGAAAC	CATTGACCTA	14400
GACCACCTCT	TTGTCAACTA	TGGTTTCAAA	TCTTCTGTCG	GTAACCTTAA	AAACTGGGGG	14460
CTCGACCTCA	ACCGTCACAA	GATTATCGTC	AACAGCAAAAC	AGGAATCCAG	CCAAGCAGGT	14520
ATCTATGCTA	TCGGTGACTG	CTGCTACTAT	GACGGAAAAA	TTGATCTGAT	TGCGACAGGC	14580
CTCGGAGAAG	CTCCAACCTGC	TGTCAACAAC	GCTATCAACT	ACATTGACCC	TGAACAAAAA	14640
GTACAACCAA	AACACTCTAC	TAGTTTATAA	AAAAGAACCA	CGAGTCACAT	AGGATTCTGT	14700
GTTTTATAAT	TCATCCGCTA	TCTTATTGAT	TTTTCTGAGT	CTGTGATTGA	CACCACTTTT	14760
GGTCAGAGGG	GTGCTGAGAC	TATCTGCTAA	CTGCTGGATA	GAGTAGTCTG	GGTGCTGAAT	14820
CCTCAGTTGC	GCCACTTCCT	GCAAATCTAC	TGGCAAATTT	TCTAAGCCCA	TGATATCTTT	14880
GATTTTACTG	ATATTGTTAA	TGGTCTTCAT	GCTGGCAGAA	ACTGTCCGAG	CGATATTAGC	14940
TGTCTCGGCA	TTATTAGCCC	GATTGAGGTC	GTTACGGGTT	TCTCGCAAAA	TCTTAACCCG	15000
CTCAAAATCA	TCACGTGCCT	GCATGGCTCC	TATTACTATC	AAGAAGTCCA	TAATGTCTTC	15060
TGCTCGCTGG	AGATAGGTCA	CAGCCCCCTT	CTTGGCTCA	AGCACCTTGG	CATCCAGTAA	15120
AAACTGTTGG	AGAAGGGAGG	CAATTCCTTG	CGCGTGGTCC	AGATAAACAG	AACTGATTTT	15180



CAACTGGTAC TTGCCTGACT CAGGGTCACG AATGCTCCCA TTTGCCAAGA AAGCGCCACA	15240
GAGATAGGCA CGACCTGCTT CCTCATCCGA TAAAATCGCC TCATCAATAC CTGTTTCCAG	15300
GCCAAAGAAA GACTCTGCCA AGTGCAAATC ACTTAACAAA TCCTGCACCT TTTTCATCTGT	15360
AAAAACGGTA TAGACGCGAT TCTTGCGAAG ATTGCTCCGT TGGTGGTGAC GAATTTTCAGA	15420
TTTGATTTC TAGAGATGGA GAAAGGACTC ATAGAGGTGA CGGGCCAGTT TGGCATTTTC	15480
TGTCACAACT GACAAAGTCA AGCCCGAAGT CGAGAGACCG ATGCTACCAG ACATTTTGAT	15540
AATGGCAGAT AATTCATGCC AGCTCAGATG GTGTTGGCCC AGGATTTCCT CTTTACTGC	15600
TACTGTGAAA CTCATTTTTT CACCTGTATA ATGCGCATCA ACTCGTCCAC AATCAAATCT	15660
CCATCGTGGG AGGCACCGCC ATTTTCCAGA CGAAGGAAGT TAGATGAAAT CACGCGCGAA	15720
ACTTGCTTAC AAAGACCTAC AAAATCGTGT TCCACTTGCA CTAAGTATTC ATCAAAACGG	15780
TTGGAATTCA TGTATTCCTG AGGCACTTTT TCAATATTCA CCAAGACAGT GTCGATAAAA	15840
GGGCGACCAA GGTGACGATG CAAGACTTCC ACGTGGTCGC TATCTGTAAA GTGTTCCGTC	15900
TCCCCACGTT GGGTCATGAT ATTGCAGACA TAGGCAATTT CTGCCTTGGT TTCCAAAAGA	15960
GCCCCGCCAA TTTCCTTAAT CACGATATTG GGCAAATAG AGGTAAAGAG GGAACCTGGC	16020
CCTAGGACAA TCATGTCACT TTCAAGGATG GTCTGCACTA CTCGACGGCT GGCCAGAGGC	16080
GTATCATCGT TTAGGGCATT GGTACATAG ACATTGTCAA TTATGCCTCG ATGGTCTACA	16140
ATATGACTCT CTCCAGCCAC TTCTGTCCCA TCCTGAAAGA CTGCATGAAG GGTCAAAGGA	16200
TGGTCACTGG AAGGATAAAT TTTCCCTGTT GTATGGAAAA ATTTGCTCAA TAACTGCATG	16260
GCATTATAGG TTGAACCTG CATTTCTGAC AAGCCAGCAA TGATGAGATT TCCCAATGGA	16320
TGGCCAGCAA AGGCTCCGGC ATCCTCAGAG AACCATACT GAAAGACCTT CTCATAAAAC	16380
TTAGGCATAT CCGACATGGC CACAAGGACA TTACGAAGAT CACCTGGCGG TGTCAACTGT	16440
TGCATATTTT TTCGGAGTTC ACCTGAAGAA CCACCATCAT CTGCCACCGT CACGATAGCT	16500
GGGATTTCCA CATCTTTTTC CCGCAGACTT TTTAGAATGA CGGGACTTCC AGTCCCTCCA	16560
CCAATCACCG TTATCTTTGG TTTTCTCATG AACGGTTTAC CGTTTCCTTT CTGCGGTCTT	16620
TGTCGCGATG CCCTTCATTA ACAGACCAAT TCTTGATAA GTCCTGCGCC AAGCGTTTAG	16680
CAAATGCCAC ACTACGGTGT TGTCCACCCG TACATCCCAT GGCAATGGTC AAAACGGACT	16740
TACCTTCCTT TTGGTAACTT GGCAGAATCG GCTCAATCAA GGCCAATAAA TGTGATAAA	16800
AGTCTTCTGA CTCAGGATGG TTCATGACAT AATCATAAAC AGGTTTCATCC ACACCCGTTT	16860
GGTTTCTCAG TTCTGGTAAA TAATAGGGAT TTGGCAAGAA ACGGACATCA AAGACCAAGT	16920

500

CCGCATCAAT	CGGGATTCCA	TACTTAAATC	CGAAAGACAT	GACTTCGATA	CGGAAAGACT	16980
GGGCTTGTTT	TTGGTCTGAA	AACTGCTCTG	CAAGGGTTT	GCGCAGCTCA	CGTGGAGTGA	17040
GTTCAGTCGT	ATCCACCACA	TTTGGCTCA	TATTTTCAA	AGGTGCCAAG	AGTTCACGTT	17100
CCAACTTGAT	TCCATCTAAA	ATACGACCGT	CTGCTGCTAG	TGGGTGACTC	CGTCTGGTTT	17160
CCTTGTAACG	AGCGACCAAT	TCCTTATCAG	CCGCATCCAA	AAAGAGGATT	TTGAAATCCA	17220
AACCATCTTG	ATTTTCCAAC	TCATCCAAAA	CAGCTTGAAT	CTCTGAAAAG	AAAGAACGGC	17280
TACGCATATC	CACTACCAAG	GCCAACTTAG	GATTGTCTTC	CTTAATTTCA	ACCAGCTGCA	17340
AAAACCTTAGG	CAAGAGAGCT	GGCGGCATA	TATCAATGGT	GAAATAACCT	AGATCCCTCGA	17400
AGGACTGAAT	GGCTACAGTT	TTCCCTGCGC	CACTCATCCC	TGTCACAATC	ACCAAGTGAA	17460
GTTGTCTCTT	TGTCATCTTT	TTCTCCTTAT	ATCAAAAGAA	GTTTGGCAAC	ACCAAACCTC	17520
AACTAGCTTA	TCCAACTCTCT	GCGATGACTT	CAATTTGAC	TTTTACATCA	CGAGGAAGAC	17580
GAGCTACCTC	CACAGCTGAA	CGAGCTGGGA	ATTCTCTTTT	GAAGGCCGTT	TGGTAAACCT	17640
CATTAAAAGG	AACAAAGTCG	TTCAATATCGC	TCAAGAAGCA	AGTTGTTTTG	ACAACATGGT	17700
CAAAGTCTGT	TCCTGCTTCT	GCCAAAATAG	CACCGATGTT	TTTCAAGACT	TGCTCTGTCT	17760
GTTCTTGAT	ATTCTCTCCT	ACAATTTCCC	CAGTTTCAGG	GGATAGGGGA	ACTTGACCGC	17820
TAGCAAACAA	AAGGTTGCCA	ACGATTTTTC	CTTGAACATA	GGGTCCGATA	GCCTTTGGGG	17880
CCTTATCTGT	ATGAATTGTT	TTTGCCATTT	TCTTTCTCTC	ACAATTTTTC	TAAGATTGCA	17940
TCCCAAGCCT	CATCCATCCC	TGCCTTACTG	ACAGATGAAA	AGAGGATGAA	ATCGTCACTC	18000
GGGTCAAAGT	TTAATTTCTT	TTTGATTGCT	GATTCATGCT	TGTTCCATTT	ACCACGAGGA	18060
ATCTTGTCGG	CCTTGGTCGC	CACAATGATG	ACTGGAATCT	CATAATACTT	GAGAAATTCG	18120
TACATCTGCA	CATCATCTGC	TGACGGGTCA	TGACGAAGGT	CAACTAGACT	GACAACCGCA	18180
CGGAGATTTT	CCCGAGTCGT	TAAGTACTCC	TCAATCATGC	ACCCCACTT	TTACAGTTCC	18240
TTTTTAGAAA	CACGAGCATA	GCCATAACCA	GGCACATCCA	CAAAGCGCAT	CTTGTCATCA	18300
ATGTTAAAAA	AGTTCAGGAG	CTGGGTTTTA	CCAGGTTTTC	CTGATGTACG	GGCGAGATTC	18360
TTACGGTTCA	ACATAGTGTT	GATAAAGCTG	GATTTACCAA	CATTTGAACG	CCCTGCTAGG	18420
GCAATCTCTG	GCAGTTCATC	CTGCGGATAG	TGGGACTTAT	TAGCTGCAC	GAGCAAGATT	18480
TCAGCATTTG	GTGTATTAAG	TTCCATAGTC	ACCTCTAGGC	TGTTTCTAGG	ATCGGTTTAT	18540
CCGTTCCATC	TACAGTTTCT	TTAGTGATGC	GAACCAATTT	CACATTTTCC	TGACTCGGCA	18600
CCTCAAACAT	GACATCTAGC	ATGGTTTCTT	CGATGATGGA	GCGAAGTCCA	CGCGCCCTTG	18660
TCTTCCGTTT	GATTGCTTTA	TTAGCAATCT	CTTGAAGGGC	TTCTCGTCA	AATTCCAACT	18720

501

CAACATCATC	ATAAGAAAGC	AAGGTTTGGT	ATTGTTTCAC	CAAGGCATTT	CTTGGCTCTT	18780
TCAAGATGCG	AACCAAGTCA	TCAACGGTCA	ATTGCTCAAG	AGCCGCAAAA	ACAGGCAAGC	18840
GTCCAATCAA	CTCAGGGATA	ATACCAAATT	TTTGAATGTC	TTCAGCGATG	ATTTCTTGCA	18900
TGTATGAGCT	GTTTTCGTCA	ATCGCCTTAT	TATTTTGACC	AAATCCGATG	ACTTTTTCAC	18960
CCAGACGTTG	TTTGACAATT	TCTTCAATAC	CATCAAAAGC	ACCACCCACG	ATGAAGAGGA	19020
TATTTTTTGT	ATCCACTTGA	ATCATCTCTT	GTTGTGGATG	TTTGCCTCCA	CCTTGAGGCG	19080
GTACGCTAGC	AACAGTTCCC	TCAATAATCT	TGAGAAGGGC	TTGTTGCACC	CCTTCACCAG	19140
AAACATCACG	TGTGATAGAC	ACATTCTCAC	TCTTCTTGGC	AACTCTGTCA	ATTTTCATCCA	19200
CATAGATAAT	GCCACGCTCT	GCACGTTTCA	TGTTAAAGTC	AGCAACCTGC	AAGAGTTTGA	19260
GGAGGATATT	TTCCACATCC	TCACCCACAT	AACCAGCCTC	CGTCAGAGCT	GTCGCATCCG	19320
CAATAGCAAA	AGGTACATTC	AAGCTCTTAG	CCAAGGTCTG	GGCAAGGAAA	GTTTTCCCTG	19380
AACCAGTTGG	GCCAATCATC	AAAATGTTTG	ACTTCTGCAA	ATCCACATCT	TCTGACTCTT	19440
CGCGTGTATC	GTGGAAATTG	ATGCGTTTGT	AGTGGTTATA	AACCGCCACT	GCCAAGGCAC	19500
GCTTGGCAGC	ATCTTGACCA	ATTACATAGT	GGTCAAGAT	ATGGAGGAGT	TCAATTGGTT	19560
TTGGCACCTC	AGACAAGTCT	GCCAAGACTT	CCTCAACCAA	TTCTTCTCGA	ATGATTTCCT	19620
GAGCTAACTC	CACGCATTCA	TTACAAATAA	AAGCATTGTT	GCCAGCAATT	ATTTTTTGTA	19680
CTTCTTCTTG	GTTTTTGCCA	CAAAATGAGC	AATAAACCAT	CATATCATTT	TTTCTATTTG	19740
TAGACATGAT	TTCTTCCAT	TCTATACTGT	CATTCTATCT	AAAATAAGGT	CATGTAAAAA	19800
GCATGAATAC	TATTGACCAG	ATTGGTAAAG	GCATTTAACC	AAAGGAGGAT	AGAAAGCCCC	19860
TAACGCTTTT	TACGAAAAGC	TTGTGCTCCT	GCCAGAAAAGC	AGATGAAACA	CAGAAAAGCC	19920
GTGAATAGAC	CAAATAAACT	CCGTTCATT	AGACTTCCTT	TCTCTTGCGG	TATTGGATGG	19980
TAAATCATA	AGGATTCTTC	TCATCTTTGG	CGTAAATTTT	GCTTGAAACT	GTCTCAAAAA	20040
GAGACAAGTC	AAGTTCTTCA	GGGAAATAGG	TATCTCCTTC	CACCCGAGCA	TGAATGTGAG	20100
TGACAATCAC	TTTATCAAGG	TAAGGTTCAA	AAGCCTGAAA	AATTTGCTTC	CCACCGATAA	20160
TGTAGAGATT	CTTTTCTTGA	GCCTGATACC	AGTCAAGAAC	AGACTGGACG	TCCTGAAAAG	20220
TAGCAACCCC	ATCTATCTTT	TCTTCCGGAT	TACGCGTCAA	AATCAAGGTT	TCCCGTTTTG	20280
GAAGCAAGCG	ACGCCCCATC	CCATCAAAGG	TCACACGCCC	CATCAAGATA	GCATGATTCA	20340
GAGTTGTTTC	TTTAAAGTGC	TGCAATTCTG	CTGGCAAATG	CCAAGGCAGA	CGATTTTCCT	20400
TACCAATCAC	ACCCTCTTCA	TCCTGGGCCC	AAATAGCTAC	GATTTTCTTA	GTCATGCTTC	20460

502

CATCCTTTTC ACTGATAGTA CTATTTTATC AAAAAACTCA AAAAAAGACT GGTTTGGAAT	20520
AGCTTACAAA ATAGAAAAAA TCTGTAAGAA ATTCCTACA GATTATCTA TGTTTCCTTA	20580
TTTCTTACAA ACCAGGTGCT TGTCCAAGTT CGGCTGCAAG CATCCAAATT GTTTTATCTG	20640
TTTCAGTTTT AGCGCTGCA AAGATACCGT TTGTCACATC GTCACCTTCT TCATCAGTGA	20700
CATCCAAACC TTTTGGAAA AGTTCTGACA AGTAACGGTA GATAACAAGA ACACGTTCCA	20760
AGCTTTCTTC AACATTACGG TATTCACCAG CTCTTCTTC GATTTCACTA TTTGAAGGA	20820
ACTCTGTCAA TGTAAGAAT GGGCTTCCAC CGAGTGTAAT CAAGCGTTCA CTGATTTCAT	20880
CCAATTGACC GTCAAGAGCT TCCATGTACT CATCCATTTT TGGATGCCAT ACAAGGAAAC	20940
CACGACCATG CATATACCAG TGCACCTGGT GCAAAGCAAC GTGAGCTACA TACAAATCAG	21000
CAACAGCTTG GTTCAAGACT TCCTTTGTTT TTGCCAATGC	21040

(2) INFORMATION FOR SEQ ID NO: 56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2387 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

ATTCTTAATA CGATTAAAG GCTTATTACT AAAAGAAAAT TTCAGTTAGA TGAACATAAC	60
TTGCTCGTCA AATCCCGATT TAACGAGATG TTTGGGGAAA ATAAATATT TGAAGCATT	120
GATAACTTAT TTGATATTAT AGATCGTGAT AGGGGCAAAA ATTATCCTAA ATCAGATGAG	180
TTGTTTAGTG AGGAGTACTG TTTATTTTAA AATACAAAGA ATGTTACTAA AAACGGATTT	240
TCATTCGATA CAAAGCAATT TATCACTAAA ACAAGGATA AATTACTTCG AAAAGGCAAA	300
CTTGAGCGTT ATGATATAGT CTTGACAACA AGAGGTACTG TTGGAAATGT AGCGTACTAC	360
GATGAATTAA TAAAATATAA ACATTTACGT ATAAATTCAG GTATGGTAA ATACGTCCC	420
AAGACACCAA ATCTAAATCA GAAATTTATT ATCCATGTTT TAAGGAATAA TAATTATAGT	480
CGAGTGATAT CAGGAAGTGC TCAGCCTCAG TTACCAATTA CAAAATTAAA AAAAATACTT	540
CTCCCCCTCC CCCCCTAGC CCTCCAAAAT GAGTTCGCAG ACTTTGTAGT CCAGGTCGAC	600
AAATCACAAT TGGCAATCCA AAAATCTCTG GAAGAACTTG AAACCTTGAA GAAATCTCTG	660
ATGCAGGAGT ATTTTGGCTG ATATTCTGCC ATTGTAATTA CGGTAATGAT TTGTTATAAT	720
ACTTCAAAGG AGGAAATCAG ATGGTAGTAA AAACAAGAAA ACAAGGAAAT TCAATCACCA	780
TTACGATTCC AAGTGAAATT AATATTCCAA GTGGTGTAA ATACGAAGCG AAATTGTTAC	840

503

CAAGTGGTGA GATTATCTTT ACTCCTGAAG AATTGGGGCA GCAGGTTTCT TATGTATCTG	900
ATGATGCCTT TGA CT TAAAT TTAGATAAAA TATTTGACGA ATACGACGAT GTTTTCAAAG	950
CTTTGGTGGG AAAATGACAA TCTATTTGAC AGAAAAGCAA ATTGAAAAAA TAAATGCTTT	1020
AGCAATTCOA CGGTATTCTC CAAATGAGAA AATTCAAACA GTTAGTCCTT CTGCCTTAAA	1090
TATGATTGTG AACTTACCAG AACAAATTTGT CTTTGGGAAG CCTCTTTATC CAACAATTTT	1140
-----	
TGATAAAGCA ACGATACTAT TTGTCCAATT GATAAAGAAG CATGTTTTTG CTAATGCTAA	1200
TAAAAGAACT GCTTTCTTCG TTTTGGTCAA ATTTTACAA TTAAACGGCT ATCGTTTTTC	1260
TGTAACGGTA GAAGAAGCAG TAAAAATGTG TGTAACCATC GCAGTAGAAG CTTTAACTGA	1320
TGAAAAAATG ACAAGCTACT CCAATGGAT TTCTGAACAT TCTGTTAGAG AAAAGGTCOA	1380
AAAGTAACCT AGTATGCTGG ATTTGAATGA GCACAAGAAA ATAAATGAAC AGACAATATT	1440
AGAATTCCTG AATGCAGAAA CTGATATTGT CTCTTTTAT TGATGAATAA GAAAGTGAGA	1500
AATTATGGAA TCAAAAGTTA CAATTATCAT GCAAGAAATG TTACCTCTTT TAAATAATGA	1560
ACAATTACTA GCGTTGAGAG AGAGTTTAGA ACATCATCTA GTAGACGGAA AAAAGCAGCA	1620
GAAGTATTCG AATAATAACC TGTGCAACT ATTTATTACC GCCAAGCAGG TAGAGGGCTG	1680
TAGCTCAAAA ACAATTCGTT ATTATCAGAG GACGATTGAA AACTTCTTTA ATGCTATTAA	1740
AGAGTCTGTG ACACAACCTA CAACAGATGA TTTAAGGAGT TATTTAGCAA ATTACCAGTC	1800
TGAAAAGGAT TGTAGTAAGG CAAATTTAGA CAATATTAGG CGTATATTGT CTCTTTTTTT	1860
TGCTTGGCTT GAGCAAGAGG ATATATCATT AAAATCCCA TTCGACGGAT ACAGAAAATT	1920
AAGACTGAGC AAAATGTGAA GGAAACTTAT ACTGATGAAC ATTTGGAAAT TATGCGTGAT	1980
AACTGTGAAA ATTTGAGAGA TTTGGCAATA ATAGACCTAC TAGCATCCAC AGGTATGCGT	2040
GTAGGGGAGC TTGTACAGTT GAATCGTTCA GATATTGATT TTGAAAACAG AGAGTGTGTT	2100
GTCTTTGGTA AAGGAAACAA GGAGAGACCA GTATATTTTG ACGCTCGTAC GAAAATTCAT	2160
TTAAGAAATT ATCTTAACGA CAGAAAAGAT AGTCACCCTG CTCTTTTGT AACGCTAGTT	2220
GGAAAAGTCC AGAGGCTTGG AATTGCTGGT GTAGAGATTC GCTTAAGAAA GTTAGGAGAC	2280
AAACTCGGCA TACAAAAGGT TCACCCACAT AAGTTCAGAA GAACTTTAGC GACTAAGGCA	2340
ATTGATAAAG GTATGCCTAT CGAACAAGTC CAAAACTGC TAGGTCA	2387

(2) INFORMATION FOR SEQ ID NO: 57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10669 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

504

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

ATATTAAAGC GACTTTCTGT GCGCTAGGGA AAAATGTTCC TGGGAATGAG GACTTGGTGA	60
AGAGGATAAA ATCTGAAGGT CATGTTGTTG GAAACCATAG CTGGAGCCAT CCGATTCTCT	120
CGCAACTCTC TCTTGATGAA GCTAAAAAGC AGATTACTGA TACTGAGGAT GTGCTAACTA	180
AAGTGCTGGG TTCTAGTTCT AAATCATGC GTCCACCTTA TGGTGCTATT ACAGATGATA	240
TTCCGAATAG CTTGGATTG AGCTTATCA TGTGGGATGT GGATAGTCTG GACTGGAAGA	300
GTAAAAATGA AGCATCTATT TTGACAGAAA TTCAGTATCA AGTAGCTAAT GGCTCTATCG	360
TTTTGATGCA TGATATTCAC AGTCCGACAG TCAATGCCTT GCCAAGGGTC ATTGAGTATT	420
TGAAAAATCA AGGTATACC TTTGTGACCA TACCAGAGAT GCTCAATACT CGCCTAAAAG	480
CTCATGAGCT GACTATAGT CGTGATGAAT AAGCAAGAAA AAATAGGTCT GTTAGATATT	540
TGACAGACTT ATTTTTTACA GAATATAGTA CTACTTAAAA AATGTTTTAT GCTATAATTG	600
ATGAATAAAA TAGAAGGAGA AGCATATGAA TACCTATCAA TTAAATAATG GAGTAGAAAT	660
TCCAGTATTG GGATTGGA CTTTAAAGGC TAAGGATGGA GAAGAAGCCT ATCGTGCAGT	720
GTTAGAAGCC TTGAAGGCTG GTTATCGTCA TATTGATACG GCGGCCATTT ATCAGAATGA	780
AGAAAGTGT TGTCAAGCAA TCAAAGATAG CGGAGTTCCA CGTGAAGAAA TGTTCGTAAC	840
TACCAAGCTT TGGAAATAGC AGCAAACCTA TGAGCAAACCT CGTCAAGCTT TGGAAAAATC	900
TATAGAAAAA CTGGGCTTGG ATTATTTGGA TTTGTATTG ATTCAATTGC CGAACCCAAA	960
ACCGCTCAGA GAAATGACG CATGGAAAAC TCGCAATGCG GAAGTTTGA GAGCGATGGA	1020
AGACCTCTAT CAAGAAGGGA AAATCCGTGC TATCGGCGTT AGCAATTTTC TCCCCATCA	1080
TTTGGATGCC TTGCTTGAAA CTGCAACTAT CGTTCCTGCG GTCAATCAAG TTCGCTTGGC	1140
GCCAGGTGTG TATCAAGATC AAGTCGTAGC TTAATGTCGT GAAAAGGGAA TTTTATTGGA	1200
AGCTTGGGGG CCTTTTGAC AAGGAGAACT GTTTGATACC AAGCAAGTCC AAGAAATAGC	1260
AGCAAATCAC GGAAATCGG TTGCTCAGAT AGCCTTGGCC TGGAGCTTGG CAGAAGGATT	1320
TTTACCACTT CAAAATCTG TCACAACCTC TCGTATTCAG GCTAATCTTG ATTGCTTTGG	1380
AATTGAACTG AGTCATGAGG AGAGAGAAAC CTTAAAAACG ATTGCTGTTT AATCGGGTGC	1440
TCCACGAGTT GATGATGTGG ATTTCTAGAA AATCATAAAA AGAATTGTAC ATTATTCTAA	1500
TTTTTGATAT AATAGTCAGC AGGAAAGAAA GTCTTATGGC GTTCTTCAAG CGAGCTTGGG	1560
ATAGTGGGAG CCAAGTAGGG CAAAATAAAG GGCTGGCGCT TTCTGTAGTA TTTTCAAAAA	1620

505

CAATGAAGTA ATAAATTAGG GTGGAACCGC GTTCTGACG CCCCTAGGTT AAATCAACCT	1680
AGGATTGTCA GATGTGGTTC TTTTGCTTAT TCAGTCTATT GTGTGAAAGA AAGGAGAGCC	1740
GTGGACAACC TTTATCTTGT AAAAGACGAT AGTCAACTAG CTACATTTCG TGATTTTGTA	1800
GTAAGAAATA CTGAAAAGTT GAAAGATTAT CAATCTTTTT TAAAGAATGA ACTTGCAGTC	1860
TGTGATTTAC CGCAAGCTGT TATTTGGTCA GATTTTAATG CTGCTACACA GATTATTAGG	1920
<hr/>	
GAAAGTGCTG TTCCAACCTA TACAAATAAT AGACGAGTGG TTATGACGCC TGATTTAGCT	1980
GTGTGGAAG AATTGTATTT GTATCAGTTG ATGGACTACG AGTGTCTGA GCAAACTCAA	2040
GCAATAGAAA GTCAGTATCA TTCTTTATCT GAAAATTTCC TCTTACAGAT TGTAGGACAT	2100
GAGTTAGCTC ATTGGTCGGA CATTTTTTAG ATGATTTTGA TGTTATGAC TCTTATATCT	2160
GGTTCGAAGA GGGGATGGTT GAATATATTA GTCGCAAGTA TTTCTTGACA GAAGAGGAAT	2220
TTCAAGCGGA AAAAATTTGT AATCAATCTC TCGTAGAACT TTTTCAGAAG AAGTATAGTT	2280
GGCATTCAAT GAATGATTTT GGTCTTCGA CTTATGATAA GAACTATGCA AGTATTTTTT	2340
ATGAATACTG GCCCAGCTTT TTGACAGTAG ATAAGTTGGT AGAAAATTTA GGTAGTGATC	2400
AAGCGTCTT AGATTCTTAT CATTTATGGG CAAATACAGA AAAAATTTT CCCTTGTTAG	2460
ATTGGTTTGT TCAGCAGAAA TTAATTGAAA AAGAAATATA AAAACTAAAG GAGTAAACAA	2520
TGTCTAAGAA ATTAACATTT CACTGCATCA GTGGCAGAGA CCTCCTTACA GTCCGGCTGC	2580
TCCACGCTCA GCACTAGAGT GCCTGAGCTA GACGCAGTAC TAACTCGTCT TGCCTCGTAT	2640
GATCGACGAG GCAGACTCGT GTCGCAAGTA ATTATTTTTT ATTAAGGAGT ATTCAATGTC	2700
TAAGAAATTA ACATTTCACT GCGTCAGTGG CAGAAACCTC CTTACAGTCG GACTGCCCTA	2760
CGCTCAGCAC TAGAGTGCCT GAGCTAGACG CAGTACTAAC TCGTCTTGCC TCGTATAATC	2820
GACGAGGCAG ACTCGTGTCT CAAGAAATTA TTTTTTATTA AGGAGTATTC AATGCTAAG	2880
AAATTAACAT TTCAAGAAAT TATTTTGACT TTGCAACAAT TTTGGAATGA CCAAGATTCT	2940
ATGCTTATGC AGGCTTATGA TAATGAAAAA GGTGCGGGGA CAATGAGTCC TTACACTTTC	3000
CTTCGTGCTA TCGGACCTGA GCCATGGAAT GCAGCTTATG TAGAGCCATC ACGTCGTCTT	3060
GCTGACGGTC GTTATGGGGA AAACCCTAAC CGTCTCTACC AACACCACCA ATTCAGGTG	3120
GTCATGAAGC CTTCTCCATC AAATATCCAA GAACCTTACC TTGAGTCTTT GGAAAAATTG	3180
GGAATCAATC CTTTGGAGCA CGATATTCGT TTTGTTGAGG ACAACTGGGA AAACCCATCA	3240
ACTGGTTCAG CTGGTCTTGG TTGGGAAGTT TGGCTTGACG GAATGGAAAT CACTCAGTTC	3300
ACTTATTTCC AACAAGTCGG TGGATTGGCA ACTGGCCCTC TGACTGCGGA AGTTACCTAT	3360

506

GGTTTGGAGC GCTTGGCTTC TTACATTCAA GAAGTAGACT CTGTCTATGA TATCGAGTGG	3420
GCTGATGGTG TAAAATACGG AGAAATCTTT ATCCAGCCTG AGTATGAGCA CTCAAAATAT	3480
TCATTTGAAA TTTCGGACCA AGAAATGTTG CTTGAAAACT TTGATAAGTT TGAAAAAGAA	3540
GCTGGTCGTG CATTAGAAGA AGGCTTGGTA CACCCTGCCT ATGACTATGT TCTCAAATGT	3600
TCACATACCT TTAATCTGCT TGACGCGCGT GGTGCCGTAT CTGTAACAGA GCGTGCAGGC	3660
TATATCGCTC GTATCCGTAA CTTGGCCCCG GTCTAGCCA AAACCTTTGT CGCAGAACGC	3720
AAACGCCTAG GCTACCCACT TTTGGATGAA GAAACAAGAG CTAAACTCCT AGCAGAAGAC	3780
GCAGAATAAA GAGAGTGACA AATTACGAAA ATGGGCGAAC AGAGTGAGCC CTGAGCCAGT	3840
TGCCGCGAGT ATGAAGGTAT CCTTAGTGAA ACTAAGGATA CTAGGCCAAA TTGGAGACTT	3900
TTGGCTCCAA TTTTAGCAAT GAAACAACGA AGTTGGTTGC TTGCGTGCCA ATCACATAAG	3960
GCAAACTGGA AAATAAAAAG ATACTTTTCG GAGAAAAAAC ATGACAAAAA ACTTATTAGT	4020
AGAACTCGGT CTTGAAGAAT TACCAGCCTA TGTGTTCAG CCAAGTGAAA AACAACTAGG	4080
CGAAAAAATG GCAGCCTTCC TCAAGGGAAA ACGCCTGTCT TTTGAAGCCA TTCAAACTTT	4140
CTCAACACCA CGTCGTTTGG CTGTTCTGT AACTGGTCTT GCAGACAAAC AGTCTGATTT	4200
AACAGAAGAT TTCAAGGGTC CAGCAAAGAA AATTGCCTTA GATAGTGATG GAAACTTCAC	4260
CAAAGCAGCT CAAGGATTTG TCCGTGGGAA AGGTTTGACT GTTGAAGATA TCGAATCCCG	4320
TGAAATCAAG GGTGAAGAAT ATGTCTATGT CACTAAGGAA GAAATTGGTC AAGCAGTTGA	4380
AGCCATTGTT CCAGGCATTG TGGATGTCTT GAAGTCACTG ACTTTCCCTG TCAGCATGCA	4440
CTGGGCGGGA AATAGCTTTG AATACATCCG CCCTGTTTAC ACTTTAACTG TTCTCTTGGA	4500
TGAGCAAGAG TTTGACTTGG ATTTCCCTGA TATCAAGGGA AGTCGTGTGA GTCGTGGCCA	4560
TCGTTTTTTG GGACAAGAAA CCAAGATTCA GTCAGCATTG AGCTATGAAG AAGACCTTCG	4620
TAAGCAGTTT GTAATCGCAG ATCCATGTGA ACGTGAGCAA ATGATTGTTG ACCAAATCAA	4680
GGAAATTGAG GCAAAACATG GTGTACGTAT CGAAATTGAT GCGGATTGTC TGAATGAAGT	4740
CTTGAATTTG GTTGAATACC CAACTGCCTT CATGGGAAGT TTTGATGCTA AATACCTTGA	4800
AGTTCCAGAA GAAGTCTTGG TGACTTCTAT GAAGGAACAC CAGCGTTACT TTGTTGTTGG	4860
TGATCAAGAT GGAAAACTCT TGCCAAACTT CATTTCTGTT CGTAACGGAA ACGCAGAGCG	4920
TTTGAAAAAT GTCATCAAAG GAAATGAAAA AGTCTTGGTA GCCCGCTTGG AAGACGGAGA	4980
ATTCCTCTGG CGTGAAGACC AAAAATTGGT GATTTCAGAT CTGTTGAAA AATTAAACAA	5040
TGTCACCTTC CATGAGAAGA TTGGTTCTCT TCGTGAACAC ATGATTGCTA CGGGTCAAAT	5100
CACTGTACTT TTGGCAGAAA AAGCTAGTTT GTCAGTGGAT GAAACAGTTG ACCTTGCTCG	5160



507

TGCAGCAGCC ATTTACAAGT TTGACTTGTT GACAGGTATG GTTGGTGAAT TTGACGAACT	5220
CCAAGGAATT ATGGGTGAAA AATACACCCT TCTTGCTGGT GAAACTCCAG CGGTGGCAGC	5280
TGCTATTCGT GAACACTACA TGCCTACATC AGCTGAAGGA GAACTTCCAG AGAGCAAGGT	5340
CGGCGCAGTT CTAGCCATTG CAGACAAATT GGATACGATT TTGAGTTTCT TCTCAGTAGG	5400
ATTGATTCCA TCAGGTTCTA ATGACCCTTA TGCCCTTCGT CGTGCAACTC AAGGTGTGGT	5460
TCGTATCTTG GATGCCTTTG GTTGGCACAT TGCTATGGAT GAGCTGATTG ATAGCCTTTA	5520
TGCATTGAAA TTTGACAGTT TGACTTATGA AAATAAAGCA GAGGTTATGG ACTTTATCAA	5580
GGCTCGTGT GATAAGATGA TGGGCTCTAC TCCAAAAGAT ATCAAGGAAG CAGTTCTTGC	5640
AGGTTCAAAC TTTGTTGTGG CAGATATGTT GGAAGCAGCA AGTGCTCTCG TAGAAGTAAG	5700
CAAGGAAGAA GATTTTAAAC CATCTGTTGA ATCACTTTCT CGTGCCCTTA ACCTGGCCGA	5760
GAAGGCAGAA GGGGTTGCTA CGGTTGATTC AGCACTATTT GAGAATGACC AAGAAAAAGC	5820
TTTGGCAGAA GCAGTAGAAA CACTCATTTT ATCAGGACCT GCAAGTCAGC AATTGAAACA	5880
ACTTTTTCGG CTTAGCCCAG TCATTGATGC TTTCTTTGAA AATACTATGG TAATGGCTGA	5940
AGATCAGGCT GTCCGTCAAA ATCGTTTGGC AATCTTGTC CAACTAACCA AGAAAGCAGC	6000
TAAGTTTGCT TGTTTTAACC AAATTAACAC TAAATAAAAT TTGATAAACG GACTTTATCT	6060
TATTACAAAG GAGAAGAAAT GGATCCGAAA AAAATTGCTC GTATCAATGA GCTTGCTAAA	6120
AAGAAAAAAA CAGAAGGCTT AACACCAGAA GAAAAAGTGG AACAAGCCAA ACTACGTGAG	6180
GAGTACATCG AAGGTTATCG CCGCGCTGTT CGTCACCACA TTGAAGGAAT CAAAATTGTG	6240
GACGAAGAAG GAAACGATGT TACACCAGAA AACTACGCC AAGTACAACG TCAAAAAGGA	6300
TTACATGGCC GTAGTCTTGA TGATCCAAAT TCATAATAAT ACTCTTCGAA AATCAAATTC	6360
AAACCACGTC AGCTTCACCT TGCCGTACTT AAGTACAGCC TGCGGCTAGC TTCCTAGTTT	6420
GCTCTTTGAT TTTCATTGAG TATATGTATT CTTTCTTTTA ACAAAGATAG ATGAAACGAT	6480
AACAAAGAGA CTAGCAGTTT GTGTTTGCTA GTCTTTTTTC GCTAAAAAAG GAACCATAAT	6540
GGTTCTTAAA AACTATCATT AGTAACTTGC ACCGGCTGTA GCGTCTGCGT CACCACCGTG	6600
GCCTCCAGCA TCCCCTGAAT CAGAAGCGCC AGAAGTAGCA TCGGCGTCTC CATGACCTCC	6660
GGCAGCAGGA GCAAATGGTC CGCTACCACC CACCAAACGT TGACCAGTCT CTTTtagGTA	6720
CCAGTCAAGC CATGGTTGGA AGTTAAAGAC GATTTCATTG ATACCAGCGT ATGATCCATC	6780
AGCATAGTAC ATTGCTTGGT AGTTGTGAGT GTTGATAACA CCTGCAGGAG AACCTGGAAC	6840
GATCGTACGG ACGTATTCTT GGTTCGCTT GCGAAGTGTT CCGATAACCC ACTCTACGTT	6900

508

CTTCATACGT	GCTGGTGGAA	GAGAACCATG	AACAGTCGAC	ATACGGCTAC	CTGATTGAGG	6960
TGGTACACGT	TTAGCGAACA	TAGTGTCTGG	ATCTTGGTGA	GCGTTGTTGT	AGTAGAGGAA	7020
TTGGTTGTTG	TCGTCAGCGT	ATGTCAATTC	AAATGGCATA	GCTTTCGAAGA	ACATATCAAT	7080
TTGGTTAACT	GTTAGGATAC	CGTGGTCCAA	TTTGACATAG	GTATCACCAG	AAACAGCACC	7140
AGTGAATGCT	GCAACTTTTT	CTACCCATTC	TGGATCGTCA	GGGTCAACTT	CTGTGATGGT	7200
TGTAGCGATT	GGTTTCCAC	AATCCAAGTC	TTCTGATTCT	ATTGGTTTGG	GTTTTTTCAA	7260
TTTCGAAACG	ACTCCTACGT	ATTTAACAAA	GTTATCTAAG	CAAGTTTCAA	GGAATTTAAC	7320
AGTGCCCTCG	TTGGTGATAT	TTCCGTTGTT	ATCAAAAAGCT	TCCTTAGCTT	TACCAAGAAG	7380
GAATTCGTTA	CCTGGAAGCG	TGTAGGCATT	AACACCTGGA	GCATCAAGGA	TTTTACGAAG	7440
GTGAACCTGA	GCACGTGATG	TTCCCTGGTC	ATAGTATGAT	GCACCCACAA	TCATAACAGG	7500
CTTGTTTTCA	AATGGATGAA	CTTCGTATGA	AAGCCATTCA	AGTACAGATT	TGAGTGAAGC	7560
TGAGATAGTG	TGGTTATGCT	CAGGAGTAGC	AATGATAACA	CCATCTGCAC	GAGTAATTTT	7620
GTTATATAAA	TAACGTAATT	GGAAACTTTC	ATCCCATTTT	TCATCTTGGT	TAAACATTGG	7680
AACCTTCGTCA	ATTTCAAGAA	CTTCTAATTC	AAATTTGAGT	TTGAAGTAGC	GACGGATAAA	7740
TTCCAAGAGC	TTACGGTTAT	ATGATTGATC	GTAGTTTGAT	CCAACAAGTC	CAACAAATTT	7800
CATTCTTTTT	GGTCTCCTAT	CTTACAAATT	TTCCCAGTCA	AAGTCTTCAG	CATCTTTGCG	7860
AAGTAATTCT	TGTGCATTAC	GTAATTTTTC	TGTGATTTTT	ACAAAGATAC	GGAAGTCATC	7920
AAAGATGGCA	TCCAATTCTT	TGATAACATC	AAGGTCAACC	AAGTCGCCAC	TTGGGTTAAA	7980
TGCTTGAAGA	GAGTGTGAGA	GCAAGAATTC	ATCTGGAAGA	ACATTTGCCT	TGATTTCAGG	9040
AGCATTCAAAG	ATTTGACGAA	GTTGCAATTG	GGCAGGAGAT	GAACCAAGCG	TACCGTAAGA	9100
AGCACCTGTA	ATCATGATTG	GTTTGTTCAA	AAGTGGGTAA	ATACCATAAG	ACAACCAAGC	9160
AAGAGCGCTC	ATCAAAACAG	CTGGAATAGA	GTGATCATAC	TCAGGAGTAC	CGATAATAAC	9220
GCCATCTGCC	TCTTCGATTT	TAGCAGCAAT	TTCCAATATT	TCAGCAGGTA	CTTGCTTGTC	9280
AGCTGGTTTG	TTGAAGACAG	GAATGGCCTT	GATTTCAACA	AGTTCAATTT	CAGCTTTGTC	9340
AGTAAAGTGT	TTTTGCATGT	ATTGAAGCAA	TTGACGGTTT	GTAGAACGTT	TTGAATTTGT	9400
TCCAACAATA	GCAATAAGTT	TTAACATGAG	ATTTCCCTTC	TCTTTTACA	TAATACAATT	9460
TTAAATTTCC	ATTGAAACAG	TTGTCTCTAT	AGAGTAGGAA	TTCTGAAGA	ACAGCTTAGG	9520
TGGCCTTCTT	TATCGATGAG	GATGACTTCG	ATGCCCTCCA	AACTTTTCGAC	TTGCCAGAGG	9580
ATAGAAGCAG	GTCTTTCTCC	AAAGAGTCGA	GTGCTCCAGA	TTTCGCCATC	GACTGATTTA	9640
TCAGAGATGA	TTGTTAGACT	CGCTAGTTCC	GTTTCAACAG	GATATCCTGT	TTGACTGTCA	9700

509

AAAATGTGAT	GGTAATCTTG	TCCATCGACG	GTCAGGTGAC	GTTCATAAAT	GCCTGAAGTC	8760
ACGACAGATT	TATTGACAAC	AGGGATGGTC	ATTAAATGAT	TTCCCCTAGG	ATTGGCTGGG	8820
TCTTGAATCC	CGATTTGCCA	TGGGTTATCC	CCTCTTGCCT	GATTTTTTCC	AATGGTCAGG	8880
ATATTCCTTC	CCAGATTGAT	CAAGGCAGAA	GTCACCCCT	CTTTCCTAAG	AAATTGGGCA	8940
ACCTTATCCG	CACTGTATCC	TTTGGCTAAA	CAACCTAGAT	CGATCTTCAT	TCCTTTCTGT	9000
TTTAAAAACA	CAGTAGAAGT	AGAAGAATCT	AACTCGATAC	CATGAGGATT	GATTAGAGGC	9060
AGCACCGATT	CAATTTCTTG	AGGCTGGGCG	ACCTTGGCAT	CTGAAAAACC	GATACGCCAG	9120
GTTTGAATTA	AGGGACCAAT	GCTGATATTG	AGGTGGCTAG	AGAGCGCTAG	GCTATGCTCT	9180
AACCCAAGTG	AAATCAGCTC	AAACAGGTCT	GGATGAACCG	TGACGGGGGC	TATTCCTGCT	9240
TGATAATTGA	TTTCCATCAA	CTCAGATTCT	TGACTATTGG	CGTTGAAGCG	STATTCAGT	9300
TCTTTGAGCA	AGTCAAAGGA	TTTTTGGAGA	AAGATATCGG	CTTGCTCATC	CACTAATGAA	9360
ATAGTGATAG	TAGTCCCAT	TAGCCGTTCA	GAATGTGAAC	GAAGAGTCAA	GCTACCAACT	9420
CCTTTCTCTT	ATAGAAAATA	AGTTGTAATA	TCAAATAATC	ATCTAAATTG	AAGCCCTTAC	9480
ATTTTATTTT	CATGTTATTA	TAATACCATA	AAGTTAGAAT	TTTCACAAAC	AAAATTTGGA	9540
AAAAGTCAAG	AAATATGCTC	ATAAAATTCA	TCAGGCTTGA	AAACAGGATA	AATGGGGAAT	9600
TATTTTTGAT	AAAAAATGCT	GAAATAATAG	TACCCCTT	GTAAACGCTA	ACGGTAAATG	9660
GTATACTACT	AAGGTAAATT	TAGAATGAAG	GCAGGAAATT	TTTATGAGTA	AAATCGTTGT	9720
AGTCGGTGCT	AACCACGCTG	GTACAGCATG	TATCAATACC	ATGTTGGATA	ATTTTGGAAA	9780
TGAGAACGAA	ATTGTTGTAT	TTGACCAAAA	CTCTAACATC	TCTTTCCTAG	GATGTGGAAT	9840
GGCTCTTTGG	ATTGCTGAAC	AAATTGACCG	TGCTGAAGGC	TTGTTCTATT	CTGATAAAGA	9900
AAAATTTGGA	GCTAAAGGTG	CTAAAGTTTA	CATGAAGTCA	CCTGTTCTTT	CAATCGACTA	9960
TGATAACAAA	GTAGTTACAG	CGGAAGTTGA	AGGAAAAGAG	CACAAAGAAT	CATACGAAJA	10020
ATTGATTTTC	GCTACAGGCT	CTACACCAAT	CTTGCCACCA	ATCGAAGGTG	TTGAAATTGT	10080
TAAAGGAAAC	CGCGAATTTA	AAGCAACTCT	TGAAAACGTA	CAATTCTGTA	AATTGTACCA	10140
AAATGCTGAA	GAAGTTATCA	ATAAACTTTC	TGACAAGAGC	CAACACCTCG	ACCGTATCGC	10200
CGTTGTTGGT	GGTGGTTACA	TCGGTGTTGA	ACTTGCTGAA	GCCTTTGAAC	GTCTTGGAAA	10260
AGAAGTTGTC	CTTGTTGATA	TCGTTGATAC	TGTCTTGAAC	GCTTACTATG	ACAAAGACTT	10320
CACACAAATG	ATGGCGAAGA	ACTTGGAAGA	TCACAACATC	CGCTTGGCTC	TAGGTCAAAC	10380
TGTTAAAGCA	ATCGAAGGTG	ACGGTAAAGT	TGAACGCTTG	ATTACTGACA	AAGAAAGCTT	10440

510

TGACGTGGAT ATGGTTATCC TTGCAGTTGG TTTCCGTCCA AACACAGCCC TTGCAGGTGG	10500
TAAGATCGAA CTCTTCCGCA ACGGTGCCTT CCTTGTAGAC AAGAAACAAG AAACATCTAT	10560
CCCAGACGTT TACGCTGTTG GTGACTGTGC GACTGTTTAT GACAATGCTC GTAAAGATAC	10620
AAGCTATATC GCTCTTGCTT CAAATGCTGT GCGCACTGGT AACGTTGGT	10669

## (2) INFORMATION FOR SEQ ID NO: 58:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7542 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

CGCGTAATA GATACTTTAT GATAGAATAA AGAACAAAGAT TGACAAGTAA GAGGAAACAT	60
TATGCAAAAT CAAACACTCA TGCAATACTT TGAATGGTAT CTGCCCCACG ACGGTCAACA	120
CTGGACGCGT CTGGCTGAAA ATGCTCCACA CCTAGCTCAT CTGGGGATCA GTCACGCTCTG	180
GATGCCACCA GCCTTCAAGG CAACCAACGA AAAAGATGTC GGCTATGGGG TCTATGACTT	240
ATTTGACTTA GGAGAGTTCA ACCAAAAAGG GACTGTCCGC ACCAAGTATG GTTTCAAAGA	300
AGACTATCTT CAAGCCATTC AAGCCCTTAA AGCACAGGGA ATTCAACCTA TGGCCGATGT	360
AGTTCTCAAC CACAAGGCTG CTGCCGATCA CAGGGAAGCC TTTCAGGTTA TCGAAGTTGA	420
TCCTGTAGAC CGTACAGTTG AACTTGGAGA ACCCTTCACC ATCAATGGCT GGACTAGTTT	480
TACCTTCGAT GGTCCGCAAG ATACCTATAA TGGCTTCCAC TGGCATTGGT ACCACTTCAC	540
CGGTACAGAC TACGATGCCA AACGCAGTAA ATCTGGGATT TATCTGATCC AAGGGGACAA	600
CAAGGGCTGG GCCAACGAGG AATTGGTCTA TAACGAAAAC GGAAACTACG ACTACCTCAT	660
GTATGCCGAC CTAGACTTTA AACATCCTGA AGTCATCCAA AACATCTATG ACTGGGCTGA	720
TTGGTTCATG GAAACGACTG GTGTAGCTGG TTTCCGTTTG GATGCCGTTA AGCATATTGA	780
CTCTTTCTTT ATGCGCAACT TCATCCGCGA TATGAAGGAA AAATACGGTG ACGATTCTTA	840
TGTTTTTGGT GAATTTTGGA ACCCAGACAA GGAAGCCAAT CTGGACTATC TCGAAAAAAC	900
GGAAGAACAC TTTGACCTTG TCGATGTTTG TCTCCACCAG AATCTCTTTG AAGCCAGTCA	960
AGCTGGCGCA AACTATGACC TTCGTGGCAT TTTCACAGAT AGCCTGGTTG AACTCAAGCC	1020
TGACAAGGCT GTGACTTTTG TCGACAACCA CGATACCCAA CGAGGACAAG CCCTTGAGTC	1080
TACCGTTGAA GAATGGTTCA AGCCAGCAGC CTATGCCCTC ATTTTGTTAC GCCAAGACGG	1140
CCTTCCATGT GTCTTTTACG GAGACTACTA TGGGATTTC A GGCCAGTATG CTCAAGAAGA	1200

511

TTTCAAAGAA ATCCTTGACC GCCTCCTAGC CATCCGAAAA GATTGCGCCT ATGGAGAACA	1260
AAATGACTAC TTTGACCATG CTAAGTGTAT CGGTTGGGTA CGTTCAGGTG CTGAAAATCA	1320
ATCCCCAATC GCAGTCCTTA TCTCAAATGA CCAAGAAAAC AGCAAGTCAA TGTGTTGTCGG	1380
TCAAGAATGG ACTAATCAAA CCTTTGTAGA TTTACTTGGT AACCACCAAG GTCAAGTTAC	1440
AATTGATGAG GAAGGTTATG GACAATCCCC TGTCTCAGCT AGATCCGTAA GTGTCTGGGC	1500
AGTCAATACC ATCTAATAGC TCATAATAAC CAAGCTAGGT CCAAGCGGAT TTGGCTTTTT	1560
TGTATTCACA AAAAGACCTA CCCAAATGGA TAGATCTTTA CTTGATTACA ATTTACCTGC	1620
TACTGCATCC AACAATTCTT GGATCTTAGG TTGGTTGCTT CCTCCTGCCA TGGCCATATC	1680
TGGTTTACCA CCACCACGTC CATCGATGAT TGGTGCTAAT TCTTTGACAA GGTTCCTGTC	1740
ATGAAGGTCT TTTGTCTTGC TTGCTACAAG GACATTGACT TTGTCACCGA TAGCGGCAAC	1800
TAGGACAAGA AGATCAGAGT AGTCTTTTGT TTTCCAGTTA TCTGCAAAAAG TACGAAGGGC	1860
ACCGGCATCG GATACAGACA CTTGACTAGC AATGTAACGA TGACCGTTGA CTTCTTAAC	1920
ATCTTTGAAG ATATCGCCTG CGGCTGCAGC TCGGGCTTTT TCTTTCAACT CAGCATTTTC	1980
TTTTTGAAGT TGACGAAGTT GTTCTTGAAG TCCTTCTACC TTGTGAGGTA CTTCTTGAC	2040
TTGAGGTGCT TTCAAGGTTG CTGCGATAGC TTTAAGAGCA TCCTCTTGTT CACGATAGGC	2100
TTCAAAGGCT TCCTTACCAG TCACTGCCAA GATACGGCGA GTTCTGAAC CGATTCTTTC	2160
TTCTTTGACA ATTTTGAAGA GACCAATCTC AGAAGTGTTG TCAACATGAG TACCACCACA	2220
AAGTTCAATA GAGTAGTCAC CGATAGTCAC GACACGAACT TCCTTGCCGT ATTTCTCACC	2280
AAAGAGGGCC ATAGCTCCCA TTTCTTTAGC AGTGTCAATA TCCGTTTCAA CTGTCTTCAC	2340
TTCAAGTGCT TCCCAAATTT TCTCGTTAAC TTGCTGTTCA ATCGCACGAA GTTCTCAGC	2400
AGTTACTGCT TGGAAGTGGG TAAAGTCAAA GCGAAGGAAT TCAACTTCGT TAAGAGATCC	2460
TGCCTGTGTT GCGTGGTTTC CAAGGATATT GTGAAGGGCA GCGTGAAGCA AATGAGTCGC	2520
AGTGTGGTTT TTCATGACAC GGTGACGGCG ATTGCTATCA ATTGCCAAGG TATATTCTTG	2580
GTTCAAGGCA AGCGGTGCAA GGACTTCAAC TGTATGAAGG GCTTGACCAT TTGGGGCTTT	2640
CTGAACATTG GTCACAGTAG CCACAACCTT ACCTGACTCA TCCAAGATTT GTCCGTAGTC	2700
AGCTACCTGT CCACCCATTT CAGCATAAAA TGACGTTTCC GCAAAGATAA GAGAGGCAGT	2760
TCCTTCTGAA ACAGCTCCTA CTTCTGCATT GTCAGCAACG ATAGCTACCA ATTTAGAAGA	2820
CAATTGGCTA GCATTGTAGT TGAAGACACT TTCTACAGTG ATGTTTTGAA GAGTTTCATT	2880
TTGCATACCC ATTGAGCCAC CCTTGACAGC TGACGCACGC GCGCGTTCTT GCTGTTCTTT	2940

512

CATGGCTGCT TCAAAACCTT CACGGTCTAC AGTCATACCA GCTTCTTCAG CGATTTCCTC	3000
AGTCAATTCA ACTGGGAACC CATAAGTATC ATAGAGTTTG AAGACATCTG AACCAGCGAT	3060
AACAGATTGA CCTTTTCTT TCAAGTCTGC TACAATGCCT TGGGCAAAGT GTTGACCTGA	3120
GTGAAGGGTA CGGGCAAATG ATTCTTCTTC GCTCTTAACG ATTTTCTCAA TAAAGTCACG	3180
TTTCTCAAGC ACTTCTGGGT AGTAGCTTTC CATGATTTTT CCAACAGTTG GAACCAATTT	3240
GTAAAGGAAA GGCTCCTTGA TACCCAATTT TTGACCATGC ATAGAAGCAC GACGGAGAAG	3300
ACGACGAAGA ACATAACCAC GACCTTCATT TCCTGGAAGG GCACCATCAC CGATAGCAAA	3360
TGAAAGAGAA CGAATGTGGT CTGCGATAAC CTTGAAGCTC ATGTTGTCGC CATCTTGGTC	3420
ATAAACCTTA CCAGACAATT TCTCGACTTC ACGGATAATC GGCATGAAGA GGTCCGTTTC	3480
AAAGTTGGTC TTAGCCCCCT GGATAACGGC CACCAAACGC TCCAAACCAG CGCCCCGATC	3540
AATGTTCTTA TGTGGCAATT CCTTGTATTC GCTACGAGGA ACAGCAGGGT CTGCGTTAAA	3600
TTGTGACAAA ACGATGTTCC AGATTTC AAT ATAACGGTCG TTTTCAATAT CTTCTGCAAG	3660
CAGGCGAAGA CCGATATTTT CTGGGTCAAA GGCTTCCCCA CGGTCAAAGA AGATTTCCTGT	3720
ATCTGGTCCA GAAGGTCCCG CACCGATTTC CCAGAAGTTG TCCTCAATTG GAATCAAGTG	3780
ACTTGGATCC ACTCCCACTT CAATCCAGCG GTTGTAAGAA TCTTTATCGT CTGGATAGTA	3840
GGTCATGTAA AGTTTTTCAG CAGGGAAATC AAACCATTCA GGGCTTGTCA AAAGCTCATA	3900
AGCCCCAAGTG ATAGCTTCGT CACGGAAGTA ATCCCCGATA GAGAAGTTCC CCAGCATTTC	3960
AAACATGGTA TGGTGACGCG CGGTCTTCCC TACGTTTTCG ATGTCGTTGG TACGGATAGC	4020
CTTTTGGGCA TTGGTAATAC GTGGATTTTC AGGGATAATG GTCCCGTCAA AGTATTTCTT	4080
AAGGGTTGCT ACCCCAGAGT TGATCCACAA AAGAGTTGGG TCATTTACAG GAACCAAAC	4140
TACTGATGGT TCTACTGAGT GACCTTTGGT CGCCCCAGAAA TCAAGCCACA TTTGGCGTAC	4200
TTGTGCACTA GATAGTTGTT TCATATTGTC TCCTTATTCA CTTGTTTAAT GTGATTGGCT	4260
TTCCAGCATT TCCACATAGT CAATCGCGAC ACAGAGGGAA ATGACTAGGT CTGCATAAGC	4320
GTCTTCAAGA ACCGTTACGG TATAGGTAGA AGTCAGATGG AAGAGTTCCT TCTTAATTTT	4380
CGCAATCAAC TGATCGCGAT CATCCAGCAA TTTGAAATTC AAATCCCAGA TATTGCCCTC	4440
GATACGAAGA CCTAGATTAT CAAACTCATA CTTATCTCGC CAGAAGGTCA ACTTCTTACG	4500
AATGACAAAA CTCGAGCCAT CCCGAAGCTG AATTTCAAAA CGAGGAAGCA AGGTCAAGAT	4560
TTCTTTACTA ATCTCACTCA CTTGTTCAAC AGCCGCATCA TAGATGGTAA AGGTTTTAGG	4620
AATCTTAAAA AATGATCCCT CCACCTGATA GGCAATTTCT CCCCTGTCAT CTTTGATAGC	4680
GAAGCGTTCTG CCTCCAAGAC GAAACTTTTG TTTGACAAGA AATGTTTCA TCAACACCTC	4740

513

CAAAAATCAA AAGACAAGCT CATATCACGA AGGGCGAAAA ACCGCGGTAC CACCTTCATT	4800
CAATGAACCT GTCAATTCTCT TGTTCCTTATG CAATTGTATG ATTGAGTAGC ATGACTTCCT	4860
AGCTTAGATG GCTCGCAGCA CCGCCATTC TCTGGACTAA GACAAGTGAA AATCAATTCT	4920
CAACTTTCTT ATTATAACGT TTTTFTAAGC TTGCGTCAAC TGGAAATGAT CTCCGTTGAA	4980
TTAGACCAAT TCCCTACATC TCTGATTACT TTTTCAGGAT ATATTTTTTC TTAGTGCCAT	5040
TTTTCTTTTT ATCCCAAATT TTCATATTAC TAAACACAGC TACTAGAATA TTTCCAAATA	5100
TAAAGGTGCC TATCACCCAA TATATGGACT CAGTTGTTAG GTATTGTGCGA TCCAAGCCAT	5160
CCTTTAAATG GAATAGTATA GCAGTTTGGT TAACAATCAT AAAGGTTGGC CAGAACTTT	5220
TTTTGAAAAA AGTAGACATT TTCATTATTT GTTGCCGCTT TCTGTAAGGT TAATACTCAA	5280
TAAAAATCAA AAAGCAAAT AGGAAGCTAG CCTCAAGCTG TACTTGAGTA CGGCAAGGCA	5340
ACGCTGACGT GGTTCGAAGA GTATAGGCTT AGTATACTAC TAGGCAAGCA AATAAACAAA	5400
TAAACAATA GAATAGAAAA AGATAGGGCT CTAAAACTG ACTTCTATT CTTAAAAACG	5460
AACCAGCTTG ACTGATTGCT CTTCTTACGT TTATCTCCTA CTTCCGATAC ATTTTAACT	5520
GTAGGAAGAG GTCGCTATAT TTCCCTGTCC ATTTATGGTC AAATTTCTCA TAACTTCTA	5580
GGTGTTCAT GGTTCACAA TCGGGATAGA AGGCCTTATC TTCCTTTGTT TCCTCTGGGA	5640
GCAATTCCTT CGCTGGTAGG TTTGGTGTG AATAGCCGAC ATACTCCGCA TTTTGGAGAG	5700
CATTTTCAGG TTTCAACATA AAGTTGATAA AGGCATAGGC TGAGTTTTGG TTTTAACTG	5760
TTTTGGGAAT GACCAATTTG TCAAACCAA GATTGCTGGC CTCTGTCGGT ACCACATAAC	5820
GTAGATTTTC ATTTTTTTCT AACATTTGGC TGGCTTCACC AGAGAAGGTC ACGCCGATTG	5880
CAACATTATT CTGAATCATA TAGCCCTTCA TCTCGTCCGC AACGATAGCC TTGATATTTG	5940
GAGTCAGTTT GTAGAGCTTA TCCACTGTCT CTTCCAAGTG CTGCAGATCC TTGGAGTTGA	6000
GGCTGTAGCC GAGGGAATTG AGTCCTAGTC CCAGCACCTC ACGCGCCCCA TCAAAGAGCA	6060
TGATAGAATT CTTATACTCC GGCTTCCAAA GGTATCCCA ATGCTCAGGC GCTTCATCTA	6120
CCATGGTTTC GTTGTAGACA ATTCTAAGG TTCCCCAGAA GTAAGGGATG GAGAATTTAT	6180
TACCTGGGTC AAAGGACTGG TTGAGAACT CTGGTCCGAT ATTTTCGATT CTTCAATTT	6240
TTGAATAATC AAGCGGAACC AAGAGGTCTT CGTCCTTCAT CTTGTTAATC ATGTATTCAC	6300
TTGGAATGGC AATATCGTAG GTCGTTCCAC CCTGCTTTAT CTTAGTGATC ATGGCTTCGT	6360
TGGAGTCAAA AGTCTCGTAC TGAAGTTGAA TTCCTGTTTC TTCTGTAAAC TGAGTCAAGA	6420
GTTCAGGATC GATATAGTCT CCCCAGTTAT AGATAACCA TTTTGTACTA TCTCGACTAT	6480

514

TGATTTTACT ATCTAAATGA GTCGCAATTC CCCACAAGAC AAGGATAATC GCTGCAATTC	6540
CTGCTAAAAA TGAATAGATT TTTTTCATGC TTGCTCCTCC TTCTCACGAG AGATAAAGTA	6600
ATAACCTACA ACTAGGATAA TACTAAAGAG AAAGACTAGA GCAGACAGGG CATTGATTTC	6660
TAAGGAAATC CCCTTGCGAG CACGAGAGTA AATCTCGACT GATAGGGTTG AAAAGCCATT	6720
TCCTGTTACA AAGAAGGTCA CGGCAAAGTC ATCTAACGAA TAGGTGAAGG CCATGAAATA	6780
ACCAGTAATG ATAGACGGAG TCAGGTAAGG AAGCATGATT TCCTTGAACA TCTGAAATTG	6840
ACTAGCTCCC AAGTCATAGG CCGCATGAAT CATGTCGCCA TTCATTTCTT TGAGTCGAGG	6900
CAAGACCATC AAGACCACGA TAGGAATGGA GAAGGCCACG TGACTAGATA GAACGGTCAA	6960
AAAGCCAAGT GAAAACTTGA GTTGGGTAAA GAGAATCAAG AAGCTAGCAC CAATCATAAC	7020
GTCAGGCGCA ACCATGAGGA TATTATTGAG TGATAGAAAG GCTTCTTGGT ATTTCTTACG	7080
AGACTGGTAG ATGTAAATGG CACCAAAAGT CCCGATAATG GTCGCTATCA AGGCTGATAG	7140
GAAGGCCAAG AAAAATGTCT GAGCCAAAAT CAGCATGAGT CTCCCATCTC CAAACATGGT	7200
TTCAAAGTGA GTCCAGCTAA AACCTGTAAA GCTATTCATA TCATCACCAG CATTAAGGC	7260
ATAGCCAATC AAGTAAAGA TAGGCAGGTA GAGGACCAGA AAGACCAGTC CCAGATAAAG	7320
GTTGGCAAAT TTTTTCATCG TTCTCTCTT TCCTTAGTCA CCCACATGGT GATGAACATG	7380
GTCAGGATGA GAATCACACC GATGGTTGAA CCCATACCAT AGTTGTCAAT GGTTAGAAAA	7440
TTCTGCTCAA TAGCCGTCCC CAAGGTGATA ACGCGTTCCC ACCAATCAAA CGGGTCAGCA	7500
TGAAGAGACT CAAACTTGGG ATAAAGACCG ACTGAACCCC GG	7542

(2) INFORMATION FOR SEQ ID NO: 59:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9223 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAACCAAAT TCCGGTATTT TAACCTATGC TGTAATACC ATGAAGTCTG TCATGACAGA	60
TCAGGTCTAT AACATTAAGG TTGAGACAGA AAATGGAAAT TATGTTGGTG AAGCTAGCCA	120
TGTTTGGTTC CTTTGGACAA ATTACTTCGC TGATAAGAAA ATCTTTGAAG AAAACAAGGA	180
CGGCTATGCC AACATTTTGA TTCTGAAAGA TGCCTCTATA TTCTCCAAAT TATCCGTCAT	240
TCCTGATTTA TTAAAAGGGG ATGTTGTGCG AAATGATAAT ATCGAGTATA TCAAAGCGCG	300
TAATATTAAA ATCTCTTCAG ATAGTGAATT GGAGTCAGAT GTTGACGGAG ATAAATCAGA	360



515

TAACCTACCT GTAGAAATCA AAGTCCTAGC TCAGCGAGTA GAAGTATTTT CAAAACCGAA	420
AGAGGATTAG TATATAGAGA AAGCCTTTT TAAGGCTTTT TGTATACTTT AAAAGATAGT	480
TCCTTTAACA ACGGACATTC CTTGCAAATA GTTTTACAAA AATAGTATAC TGGATTTCATT	540
GAGTTTGAAA ACGTTTGCGT AAAATTGAA TGAATACTTT AGGAGACAAA TTGATGGAAT	600
TGAGTGCTAT TTACCATAGG CCTGAGTCGG AGTATGACTA TCTTTATAAG GATAAGAAAC	660
TCCATATTCG AATTCGAACT AAGAAAGGGG ACATTGAAAG CATCAACTTG CACTATGGGG	720
ACCCCTTTAT CTTTATGGAG GAGTTTATC AGGATACAAA AGAAATGGTC AAGATAACTT	780
CTGGTACCTT ATTTGACCAT TGGCAGGTTG AAGTGTCACT TGACTTTGCA CGTATCCAGT	840
ATCTCTTTGA GCTCAGAGAT ACAGAAGGTC AAAATATTTT GTATGGCGAT AAAGGGTGTG	900
TGGAATAATC TCTAGAAAAT CTTTCATGCA TTGGGAATGG ATTTAAGTTG CCTTAGCTTC	960
ATGAGATTGA TGCCTGCAAG GTTCTGACT GGGTTTCAA TACGGTATGG TATCAGATAT	1020
TTCTGAAAG ATTTGCCAAT GGCAATGCTC TATTAAACCC AGAAGGGACT TTAGACTGGG	1080
ATTCATCTGT CACACCTAAG AGCGATGATT TCTTGGTGG TGATTTACAG GGGATTATTG	1140
ATCATATGAA TTAATTGCAA GACTTGGGTA TTAAGGACT ATATCTTTGT CCCATCTTTG	1200
AATCTACAAG CAATCACAAG TACAATACGA CAGATTACTT TGAAATTGAC CGTCATTTTG	1260
GAGACAAGGA GACCTTTCGG GAACTGGTGG ATCAAGCGCA TCATCGTGGC ATGAAAGTCA	1320
TGCTGGATGC GGTATTTAAT CATATTGGTT CGCAATCTCT TCAATGGAAA AATGTCCTCA	1380
AAAATGGTGA ACAGTCTGCT TATAAGGATT GGTTCATAT TCAACAATTC CCAGTGACAA	1440
CTGAAAAGCT AGTTAATAAG AGAGACTTAC CCTATCATGT TTTTGGTTTC GAGGACTATA	1500
TGCCTAAGCT AAATACAGCC AATCCAGAGG TCAAGAATTA TCTTTTAAAG GTTGCGACTT	1560
ATTGGATTGA AGAGTTTAAT ATCGATGCTT GGCCTTTGGA TGTGGCTAAT GAGATTGACC	1620
ATCAGTTCTG GAAGGATTTT CGTAAGGCAG TTTTAGCTAA AAATCCTGAT CTTTATATCC	1680
TAGGAGAAGT CTGGCATACT TCTCAGCCTT GGCTAAATGG AGATGAGTTC CATGCCGTCA	1740
TGAATTATCC TTTATCTGAT AGTATCAAGG ACTATTCTT ACGAGGAATT AAGAAGACAG	1800
ACCAGTTCAT CGATGAAATC AATGGAGAGT CTATGTATTA CAAGCAGCAG ATTTAGAGG	1860
TCATGTTTAA TCTCTTGGAT TCACATGATA CAGAGCGAAT CCTGTGGACG GCCAATGAAG	1920
ATGTTCAACT GGTAAATCA GCCTTAGCCT TTCCTTTTTT AAAAAAGGA ACACCGTGCA	1980
TTTATTACGG AACCGAGCTA GCCTTGACTG GAGGACCAGA TCCAGATTGT CGTCGTTGTA	2040
TGCCTTGGGA ACGTGTATCA AGTGACAATG ATATGCTGAA CTTTATGAAG AGGCTGATTA	2100

516

AAATTCGGAA ATACGCGTCA GTAATCATTT CGCATGGCAA GTATAGCCTT CAAGAAATCA	2160
ACTCTGATCT AGTAGCTCTG GAATGGAAAT ACGAAGGACG GATCCTCAAA GCAATATTCA	2220
ACCAATCAAC AGAAGATTAT CTTTATAGAGA AAGAAGCAGT AGCACTAGCA AGCAATTGCC	2280
AAGAATTGGA TAATCAGCTT GTCATCTCTC CAGATGGATT TATGATTTTC TAAAACTAG	2340
TTGATGAAGA TTATGGTACA TTTCATACCT TATATAGTAT AATAAGGCTA GTTACTAAAC	2400
TTGTAAAGGA GAACTTAAAT GAATTGTAGA GGACATGAAA CAAGACAAAG AATTGTTAGA	2460
GATTTTGAAG TTCAGCCTAA AGCACATATT AAGCTGTTAG CAAATCAACA AAAACATAGT	2520
GATGCAGGAG CAACTATTGA AGATGAATAT TATGTATTTA TCGCTGAGAG TAAAATTGAT	2580
GGCAAGAAGG AAGTTATTCA GTGTTGCATG GGTGCGGCAA GGGATTTTTT AGAACTAATT	2640
AATCACAAAG GGCTACCTCT TTTTAATCCG CTTGTAGGTG ATTCTCATGT AAATAATAGA	2700
CAAGAATATG ACAATACAGG GAGTGGAAAT TTATAACCTG AAAAGTGGAA TGAAACTGCA	2760
AAGCAGCTTT ATAATGCTAT AATGTGCTTG ATTATTTTAT GGAATGCTAA GCCGGATACA	2820
CCTTTATTTA ATTTTAAAGA CGAAGTAATT AAGTATAAAA CATATGAGCC TTTTGAAAGC	2880
AGTATAAAAA GAGTAAATAC TACTATAAAG AATGGTAGTA AAGGGAAAAC TCTGACTGAG	2940
ATGATTAATG GCTACAGAGC GGATAACGAT ATTAGAGATG AAATTTGTAA CTTTAATATT	3000
CTGAAAAATA AAATTCGTGA TATGAAAAAC CAACAAGGAA ATACAATGGA ATCTTACTTT	3060
TAGTTATTGT TGAATTTTGG GTATTCTATA AAATATCCTA ATTGAGATTT AAATAGTAGA	3120
CTATACAATA TAGTTAAAT ATCAGTAAAA ACAACACTTT ATTGAGGTAT TGGATACGCT	3180
TTGCTAATAG CCTAATAATC ACATGTGGAG TGTGTCTACA ACGAAAAAGG TGATAATCCT	3240
TGATTTCAAG CTATTTTATA AGCATTTTGT CTTTGTAGAT AAAGGCAATT TTGACAATAA	3300
AAATCCTAAA AGGTGAATCG TTATAGATGT ATTTGTAGAT ATCGTTTGCG CATCGAAAAA	3360
ATTAATACAA GAATAAATAT TTATAGCTCT TTAGGTGACT TTTATAGAAG TAAAGTTTAC	3420
GATAGAAAAA CAAGAAATAA CCCACCATTT TTGGTGCCTT ATGCTTTTTT ATGCTATAAT	3480
GGATTTATAA AAATAAAGGA GTTTGCTATG ATTGGAAGA ACATAAAATC CTTGCGTAAA	3540
ACACATGACT TAACACAACCT CGAATTTGCA CGGATTGTAG GTATTTACG AAATAGTCTG	3600
AGTCGTTATG AAAATGGAAC GAGTTCAGTC TCTACCGAAT TAATAGACAT CATTTGTCAG	3660
AAGTTTAATG TATCTTATGT CGATATTGTA GGAGAAGATA AAATGCTCAA TCCTGTTGAA	3720
GATTATGAAT TGACTTTAAA AATTGAAATT GTGAAAGAAA GAGGTGCTAA TCTATTATCT	3780
CGACTCTATC GTTATCAAGA TAGTCAGGGA ATTAGCATTG ATGATGAGTC TAATCCTTGG	3840
ATTTTAATGA GTGATGATCT ATCTGATTTG ATTCATACGA ATATCTATCT AGTAGAAACT	3900

517

TTTGATGAAA TAGAGAGATA TAGTGGCTAT TTGGATGGAA TTGAACGTAT GTTAGAGATA	3960
TCTGAAAAAC GGATGGTGGC CTAATGGAAA TCCAAGATTA TACTGATAGT GAATTCAAAC	4020
ATGCTTTAGC AAGGAATCTT CGTTCACCTGA CAAGAGGAAA AAAGTCCAGT AAGCAACCTA	4080
TAGCGATTTT GCTTGGAGGG CAAAGTGGTG CCGGTAAGAC TACAATTCAT CGTATTAAAC	4140
AGAAAGAATT TCAAGGAAAT ATTGTTATCA TAGATGGTGA TAGTTTTCTG TCTCAGCATC	4200
CACACTATTT AGAACTGCAG CAAGAATATG GCAAAGACAG TGTAGAATAT ACCAAAGATT	4260
TTGCAGGAAA AATGGTAGAG TCTTTAGTAA CAAAATTGAG TAGTTTGAGA TACAATCTTT	4320
TGATAGAGGG AACTTTACGA ACAGTTGATG TTCCAAAGAA AACAGCACAA CTCTTGAAAA	4380
ATAAGGGATA TGAAGTACAA TTGGCCTTAA TTGCGACAAA GCCTGAATTG TCGTATCTAA	4440
GTACTCTTAT CCGTTATGAA GAACTGTACA TTATCAATCC AAATCAAGCA CGCGCAACTC	4500
CAAAAGAACA TCATGATTTT ATTGTAAATC ATCTAGTTGA TAACACACGA AAATTGGAAG	4560
AACTAGCTAT CTTTGAAAGA ATTCAAATTT ACCAACGAGA TAGAAGTTGT GTATATGATT	4620
CAAAAGAAAA TACAACCTCA GCAGCAGATG TTCTTCAAGA GTTACTCTTT GGGGAGTGGA	4680
GTCAGCTAGA GAAGGAGATG TTGCAGGTGG GGGAAAAGAG ACTTAATGAA TTACTTGAAA	4740
AATAACAAT TGATATTTT AGGAGAATAG AAATGAGAGG GTTTAATAAC AAGATAAAGT	4800
CTGTTTATCA AGAACTAACA AATTCCAAAG AGAAATTCGG TAGCTTTCAC AAGACTTTAA	4860
TTCAATTTGCA TACACCTGTT TCTTATGATT ACAAGCTATT TTCTAATTGG ACTGCAACGA	4920
AATATAGAAA AATTACTGAA GATGAACTAT ATGATATATT TTTTGAAAAT AAGAAAATAA	4980
AAGTTGATAA GACAATTTTT TTTAGTAATT TTGATAAGGT TGTTTTTCT AGTTCAAAAG	5040
AATATATTAC TTTTCTTATG TTAGCAGAGG CAATCATAAA AAATGGAATA GAAATAGTTG	5100
TAGTAACTGA TCATAATACT ACCAAAGGTA TTAATAAGTT ACAAATGGCA GTCTCAATCA	5160
TAATGAAAAA TTATCCGATT TATGATATAC ATCTCATAT TTACATGGA GTAGAAATTA	5220
GTGCAGCAGA TAAATTGCAT ATTGTATGTA TATATGATTA TGAACAAGAA TCATCGGTTA	5280
ATCAATGGTT AAGTGAAAAT ATTATAAGTG AGAAAGATGG AAGTTATCAA CATTCACCTGA	5340
CTATAATGAA GGATTTCAAT AATCAAAAAA TAGTTAACTA TATTGCTCAT TTCAATAGTT	5400
ATGACATTTT GAAAAAAGGT TCTCACTTAT CAGGTGCATA TAAACGAAAA ATTTTTTCTA	5460
AAGAAAATAC ACGATTTTGG AGTTTAATAT TAACTCGAAA GAATCTTCGC AACAACCTGA	5520
TATTCTCTAT AAAGAAGTTG GTGTATTAAG TTGGGACAA AAAGTTGTAG CCATGCTTGA	5580
TTTTTTATTA GCATATAGTG ATTATTCTAA AGACTTCAGA CCATTGATTA TTGATCAGCC	5640

518

TGAAGACAAT CTAGACAATC GTTATATTTA CAGGCATTTA GTTCAGCAGT TTAGAGATGT	5700
GAAAGCTCAA CGTCAAATTA TTTTAGCAAC ACATAATGCT ACAATTGTAA CAAATTCTAT	5760
GACAGATCAA GTTGTTATTA TGGAGTCAGA TGGAGTTAAC GGATGGATTG AATCACAGGG	5820
ATATGTTAGT GAAAAATATA TAAAAAATCA TATCATCAAT CAATTAGAGG GAGGAAAAGA	5880
TTCCCTTCAAG CATAAAATCT CTATATATGA GACGGCTTTA TCAGAGTAGA GTCAGAAAAA	5940
GTAGGTTAGA AATTTAGCCT ACTTTTTTCT TTGTCGGACA GGCATAGTGT ACATCTGAGG	6000
TCCAAGTCCT CTGTGGATAT TTGCTGCAGA TGAAACCAAT AGCGACTCCT AAGCCTGAAT	6060
ATCGTGAGGT AGGGGGGATA GGAAGGAATT AGCGAAATCA AGGTTCTACA AACAGAATCG	6120
TGACTTGAAG CCATATATAG CGGATGAGGA ACTCTAAAAT CCAAATAGGT GTCGTAACCT	6180
ATATACGTAA ATTACGAGAG TAAACTAGGA AAGATGTACG GCTTATTCCG TGAGCGTTTA	6240
GGACGTAGTA CAACGAATCA TGGGAGTCAG CTGAACACAT AGTATTGAAG AAATTCTCTGT	6300
AATGGAATG GAGCGAAGAA GTGAACAATT AAATGAATAC CTCTCTAATT AAATTGTCA	6360
ATTCTAATTC CTGGTATGAA AAGACAGTGA CCTGAAAATG TAAACGATGG GAGCTGATCA	6420
TAAATATAGG ACGGTACATG CAGTGGTGT AGAGATTAGT CCTTACTTGA TTTGTGATAA	6480
CTTCCCCAAA TTTCTTCTGC TATACTTTTC TCAACTTTTA AAAATCCAAC TAAGAATTTT	6540
ACCTGGGGGT TTGGGGGCGG AGCACTAAGT TATCTTATCG TTAGCTGTCA AAAGTGGTAG	6600
GTTTTGATAG GCTGGCGATA TGATTTTTGG GATATTGTGG ACACAATATC TGAGCTCGCA	6660
AAGCCTTACA AGAATGAAAA TCAGTTGTG GAAAAGTCTA CTGACATTGT ATGGTAGCTC	6720
ACATTGTCAG TACAAGTATT TTGGAAAGGA AGTAGCAGTA TGAAACGAGA TGTGCGTGAT	6780
ATTCCGAAAC AATTCCTTT AACAGAAGCA GAAGAAAAGC AAATTCTAGC TTTGATGAGA	6840
GAGCGGGGAG AGACTAATTT CTCTGATTTT CTTCGTAAAA GTTTACTTTC CTCTGATTTA	6900
CAAAAACAGA TGGAGACATG GTTTGCCCTC TGGCAATCCC AAAAAGTAGA ACAAATCAGT	6960
CGTGACGTTT ATGAAGTTTT AATCTTGGCA CAGTCAGAAC GTCAAGTCAC CCAAGAGCAT	7020
GATCTATTC TCTTAACGTG CGTGCAGGAA TTGATTCAAG AGGTTGCAAA CACCATACCC	7080
CTCAGTAAAG AATTCGTGA GAAGTACATG AGGTAAGCAC ATGGAACATC GTTACCGAAC	7140
CAATCTCAAG AAAGTGTTTT TGTCTGATAG TGAGTTGAAC CAACTAAATA TAAATATCGA	7200
TCAAAGTGGT TGTAATCCT TTTCTGAATA TGGGAGACGA ACTCTACTCG ATCCTGGTAT	7260
GAATTTTATC ACGATTGACA CAAACGGTTA CCAAGATTTA GTGTTTGAGT TAAAGAGGAT	7320
TGGCAATAAT ATCAACCAGA TTGCTCGAAG TGTTAATCAA TCTCAGTTAA TTTCTGGTGA	7380
AGAATTGCAG GAGTTGAAAA AAGGAATTGG TGAATTGATA AAAGAAGTTG ATAAGGAATT	7440

519

TAATCTGCAA GCGCAGAAGC TAAAGGAGTT CCATGGTCAT CACTAAACAC TTTGCCATTC	7500
ACGGAAAGAG TTACCGCAGA AAGCTTATCA AGTACATTCT CAATCCTGAG AAAACCAATA	7560
ATCTTGCCCTT GGTGTCGGAC TATGGCATGA AGAATTTTCT GGACTTTTCT AGCTATGAGG	7620
AAATGGTGCA GATGTATCAT GAAAATTTCA TCAGCAACGA TACGCTTTAC GATTTTCGCC	7680
ACGACAGGAT GGAAGAAAAT CAACGAAAAA TACACGCTCA CCACATCATT CAGTCTTTCT	7740
CGCCAGAGGA TCATATCACT CCTGAACAAA TCAATCGGAT AGGTTATGAG ACTGTGAAGG	7800
AATTAACTGG TGGCAAATTT CGTTTTATCG TTGCGACCCA TGTTGATAAA GACCACCTGC	7860
ACAATCACAT CATTATCAAT TCAGTAGATA GCAATTCTGA CAAAAAGCTC AAGTGGGACT	7920
ACAAGGTGGA GCGAAATCTT CGCATGATTT CTGACCGTTT TTCTAAAATC GCAGGTGCTA	7980
AAATCATTGA GAACCGCTAT TCTCACCAGC GGTATGAAGT CTATCGTAAG ACTAATCACA	8040
AGTATGAACT CAAGCAGCGA CTCTATTTTT TGATGGAACA TTCTAGGGAC TTTGAGGATT	8100
TCAAAAAGAA TGCTCCGCTA CTACATGTGG AGATGGATTT CCGTCACAAG CATGCCACCT	8160
TTTTTATTAC GGACTCAACT ATGAAACAGG TGGTGCCTGG CAAGCAACTC AATCGCAAGC	8220
AGCCTTACAC AGAAGAATTT TTTAAGAACT ACTTTGCCAA AAGAGAAATA GAAAGTCTCA	8280
TGGAATTTTT ATTGCTGAAA GTTGAGAATA TGGATGATTT ACTTCAGAAA GCAAAACTTT	8340
TTGGACTAAC TATCAATCCT AAACAAAAGC ATGTTTCTTT TCAATTTGCA GGAGTGGAGG	8400
TAAAGGAGAC AGAGCTAGAC CAGAAAAATC TTTATGATGT AGAGTTTTTC CAAGATTATT	8460
TTAAAAATAG AAAAGATTGG CAAGCTCCAG AAAGTGAGGA TTTCGTTCAA CTTTATCAAG	8520
AAGAAAAGTT ATCCAAAGAA AAAGAACTTC CAAGCGATGA GAAGTTCTGG GAGTCCTATC	8580
AAGAGTTCAA GAGTAACAGA GATGCCCTTC ATGAATTTGA GGTGGAGTTG TCACTCAATC	8640
AAATTGAAAA AGTAGTGGAT GATGGAATTT ACGTCAAGGT CAAGTTTGGT ATTCGTCAGG	8700
AGGGACTTAT CTTTGTGCCG AACATGCAGC TTGATATGGA AGAGGATAAG GTGAAGGTTT	8760
TCATCAGGGA AACCAGCTCC TACTATGTCT ACCACAAAGA CGCTGCCGAG AAAAATTGTT	8820
ATATGAAAGG TCGAACCTTA ATTAGACAGT TCAGCTATGA AAATCAAACC ATTCCATTAC	8880
GCAGAAAAGC GACAGTCGAT ATGATTAAAG AGAAGATTGC GGAAGTGGAT GCTTTGATTG	8940
AACTGGAAGT AGAAAATCAA TCTTATGTCA CGATTAAAGA TGAGTTAGTG CATGAACTAG	9000
CAGCGTCTGA ATTGAGAATC AATGAGTTGC AAGAACGAAT GTCAACCTTG AATCAAGTAG	9060
CAGAATATCT ACTGGCTTCA GTTGAAAGTA AGCAAGAAAT GAAATTAAAT CTTTCAAAAC	9120
TGAATATAAC TGAGAATATC AGTGCTAATA TTGTTGAGAA AAAATTGAAG AGCCTGGGGA	9180

520

ATCAACTGGA ATTGGAAAGG GGCAGGTATG AAAAGATGGT AGT

9223

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6827 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

TCTGCTGGCT ACCATCATCT GACTTGGGCA AGACCAAAGT CTTAGTTACA ACTGTATTCT	60
TCTCAGCATT TTCAATAACT GGCAATGCCG ACTGAAGCGT ATCTTTTCTT GTTTTTGTAG	120
CTGGTCCAGT TTCTTTTTC TGTCCGAAC CAACCAGGAC AAAAAGGAAA GCTAGACTAA	180
CAAGAACTAT TTTTTCATT TCTTCTTCT TTCTTTTGA AATTAAAATA GAATAAGACT	240
GGGAAGTGCT CCCAGCCTTG ATGTTTATAG AGCTGCACGC AAACGTGCTT CTGCATTTTC	300
TACATTACGG ACAGAGCGTG GTAGGAAGGC ACGAATATCG TCTTCCTTGT AGCCAACTTG	360
CAGGCGTTTT TCATCTACAA GGATTGGGCT CTTTAAAATT CTCGGTCTTT CCATAATCAG	420
ATTGAGAACT TCATTGACAC TCAAATCTTC AATATCCACT CCAAGGGCTT TGGCATAGCG	480
ATTTTGTAGAC GAAACGATGC TGGCTATTCC GTTATCTGTT TTGGTTAGAA TATCCAGTAA	540
TTCTTCTCTC GTAATTCCTT CTTTACCAAG GTTTTGTCTT TTATAACTTA ACTGGTGGGC	600
ATTGAGCCAG GTTTTGTCTT TTTTACAGCT AGTACAACCT GAGACTGTAT AAATTTTAAT	660
CATGTACCTA CCCCTTTCGC TACATGTTAC TATCAGTTTA GTCTATTATA CCATAAAAAA	720
CATCCGACTT GCGACCTATT TTTAATTTT TTTGACTTTT TTCGTCAATT TCGTACTTTT	780
TTCTTGACAA ACAACTAAAT GACTATCAAC TCTTTGGAG CTAGGGTCAA TAATTCACAA	840
CCTGTCTCTG TAATCAGGAT ATCATCCTCG ATACGAACGC CATATTTGCC TTCCGATATAG	900
ATACCTGGTT CATCGGTCAA GGCCATACCT GTCTTAATAG TTTCTGTAGA AGTCTGACTA	960
AAGTAGGGTT CCTCATGGAT ATCCAGACCA ATACCSTGGC CAATGCCGTG AGTAAAGTAG	1020
TCACCATAAC CTGCCTCAAT GATAATATCA CGAGGGATT TGTCAAAGTC ACGGAAACCT	1080
AAGCCTGCCT TAGCTTGGTC AATCAAGGCT TGGTTAGCTT TTAGAACCGT ATTGTAAATC	1140
TCTGCCTGCT CATCGCTAAC ATGCCCTAGA TAGATAGTCC GGGTCATATC ACTGACATAG	1200
TGGTCATAGA GACAGCCGAA GTCCATGGTG ATGGCTTCTC CCAACTCCAC TGGTTTGTGC	1260
ATTGGATGGG CATGGGGTTT AGAAGAATTG ATACCGCTAG CTAGGATCGT ATCAAAAGAT	1320
AAGCCAGATG CTCCCAACTC ACGCATGCGG AATCAAGGA AGTTGGCAAT CTCAATTTCA	1380

521

GT'TTTTCCTG GTTTGATAAA GTCAAGCGCA TCGCGGAAAG CTTGGTCTGA GATAGAACAA	1440
GCCTTGCGAA TCGCTGCAAT CTCTGCCTCA TCCTTAATCA TACGAAGACC TTCCACAAAC	1500
TGAGTTTGTG GAAGCAAGTT CAAACCTGCA AAAGCTGCCT GCATACGGTG GTAATAAGAC	1560
ACTGAAATCT CATCTTCAAA ACCGATACGA GTCAAGCCCA TGTCCTTAAC AATTCCTGCA	1620
ATGACAGCCA ATTCATCAG ATCAGCCACA ATCTCAAAAC CACTGGTTTC TTGCTTAGCT	1680
GCGATGATAT AGCGAGAGTC TGTCACTAAG ACCTGACGGT CAGCACTGAT AAAGACTGTT	1740
CCGTTTGAGC CCCAAAAACC AGTCAAATAA TAGACGTTTT TAAGATTGTT GATGATGATA	1800
CCATCTAGTT CTTTTTCTTG CATTTTAGCT AGAAATGCTT GTACGCGTTT ATTCATGATG	1860
TAAC'TTTCCT TTCAAATAGT GTCCTGTATA GCTGGCTTCG TTGGCAGCTA CTTCTTCTGG	1920
AGTTCCTGTT ACGATGATGG TTCCACCACC GACACCGCCC TCAGGTCCCA AGTCAATGAT	1980
ATGGTCTGCC GTCTTGATAA CATCCAGATT GTGCTCGATG ACGAGGACTG TATTGCCATC	2040
GTCTACAAAG CGAGCTAAAA CCTTGAGCAG GCGAGCAATG TCCTCTGTAT GAAGCCCTGT	2100
CGTCGGCTCA TCCAGAATGT AGAAAGATTT TCCTGTGAT CGTTTGTGGA GTTCGCTAGC	2160
TAAC'TTCATA CGTTGGGCTT CTCCCCCAGA AAGGGTGGTA GCTGGCTGTC CCAAGGTCAC	2220
ATAGCCTAGC CCTACATCCT TGATGGTCTG GAGTTTGGCT TGAATTTTCG GAATGTGTTG	2280
GAAAAATCTT ACCGCATCGT TGACCGTCAT ATCCAAGACC TGCGAAATAT TCTTTTCCTT	2340
GTAGTGAAC TCTAGGGTTT CACTGTTATA GCGGGTTCCG TGGCAAAC'TT CACAAGCCAC	2400
ATAAACATCT GGCAAGAAGT GCATCTCAAT CTTGATAATC CCGTCACCTG AGCAAGCTTC	2460
ACAGCGACCT CCCTTGACGT TGAAACTGAA GCGCCCTTC TTGTAGCCTC GAATCTTGGC	2520
TTCA'TTCTC TGAGCAAAAA GGTCACTAT ATCGTCAAAA ACTCCTGTAT AGGTAGCTGG	2580
GTTAGACCTC GCGCTCCGTC CGATAGGGCT CTGGTCAATA TCAATCAAAC GGTGACATG	2640
CTCAATCCCT GTAATAGTCT TAAACTTACC AGGTTTGTCT GAATTACGGT TGAGCTTCTG	2700
GGCAATGGCT TTTT'TGAGAA TGCTGTTGAT TAGAGTCGAT TTCCCTGAAC CCGACACACC	2760
TGTCAC'TCG ATAAATTTTC CTAGTGGAAA GCGAGCCGTG ACATTTTGCA AGTTGTTCTC	2820
ACGCGCTCCT ATCACTTCAA TAAAACGACC ATTTCCGACA CGGCGCTCTT CTGGTACTGG	2880
GATGACACGT TTGCCTGACA AGTACTGACC TGTGATAGAC TTGCTGTTGC GAGCCACTTG	2940
CTTAGGTGTA CCTGCTGCAA CAATCTCACC ACCAAAAACA CCGGCACCAG GACCAACGTC	3000
AATCAGATAA TCAGCCTCAC GCATGGTATC TTCGTCTGT TCCACCACGA TAAGAGTATT	3060
GCCCAAGTCA CGCATCTTTT TCAGACTGGC AATCAGGCGA TCATTGTCCC TCTGGTGAAG	3120

522

ACCGATTGAC	GGCTCGTCTA	GGATATAGAG	GACACCTGAT	AGGTTGGAAC	CAATCTGGGT	3180
TGCCAAACGA	ATGCGCTGAC	TTTCCCCACC	TGAAAGGGTT	CCTGCTGAAC	GTGACAGGGT	3240
TAGATAGTTA	AGACCCACAT	TATTAAGGAA	GGTCAAACGA	TCCTTGATTT	CCTTGAGAAT	3300
GGGACGAGCA	ATGATGGCTT	CATTTTCAGA	CAAAGTTAAC	TGGCTCACCA	AGTCCAAGTG	3360
GTCAGCGATA	GACAGGTCTG	AGATTTCTCC	AATATGTGGC	CCTTGCTGGC	CGCCACACCG	3420
GACAGACAAG	GCCTGGTCAT	TGAGACGATA	GCCTTGACAG	GTTCCGCAGG	TCAGCTCATT	3480
CATGTAGAGA	CGCATCTGAG	TGCGAGTGTA	ATCGCTATTG	GTTTCATGGT	AACGACGTTT	3540
GATATTATTG	ATAACTCCCT	CAAACGGAAT	GTGATATCG	CGCACGCCAC	CAAATTCATT	3600
CTCATAGTGG	AAATGGAATT	CCTTACCATC	TGACCCATAG	AGAATCAAGT	TCTTATCTTC	3660
TTCTGACAGG	TCCTCAAAAG	GCTTATCCAT	AGCCACTCCA	AAGACTTTCA	TGGCCTGCTC	3720
TAACATGTTT	GGATAGTAGT	TGGATGAGAT	AGGATTCCAA	GGTGCTAGCG	CTCCCTCACG	3780
TAAGGTTTTG	CTAGCATCTG	GCACTACCAA	ATCAGTATCC	ACCTCCAGCT	TGATGCCCAA	3840
GCCGTACAC	TCACTACAAG	AGCCAAAAGG	AGCATTGAAA	GAAAAGAGAC	GAGGCTCTAA	3900
CTCTGGGACA	GTAAAACCAC	AAACTGGACA	GGCATAATGC	TCAGAGAACA	ACAACTCCGA	3960
GTCGTCCATG	GTGTCGATAA	TGACATAACC	TTCTGCAATA	CGAAGGGCAG	CCTCAATGGA	4020
ATCAAAGAGA	CGACTACGAA	TGCCCTCCTT	GATAACAATA	CGGTCAACCA	CGACATCGAT	4080
ATTGTGTTGC	TTGCTCTTAG	ACAACTCTGG	CACCTCGGTC	ACATCATAGA	CTTCCCCATC	4140
CACACGGACA	CGAACATACC	CGTCTTTCTG	AACCTTCTCG	ATAAACTCT	TATGTTGGCC	4200
TTTTTTCTTG	CGGATGACAG	GAGCCAAGAT	CTGCAAGCGC	TGGCGTTCAG	GTAACCCCAA	4260
AACCTTATCA	ACGATTGCT	CCACAGAAGA	AGCATTGATA	GCTCCATGTC	CGTTGATACA	4320
GTAAGGCGTC	CCCACACGTG	CGTAGAGGAG	ACGCAGATAG	TCATTGATTT	CAGTCGTCGT	4380
TCCCACCGTC	GAGCGAGGAT	TTTTACTAGT	CGTTTTCTGG	TCGATGGAAA	TAGCTGGGCT	4440
GAGACCATCA	ATGGCATCTA	CATCTGGTTT	TTCCATATTT	CCCAAGAACT	GACGAGCGTA	4500
GGCGGACAAA	CTCTCTACAT	AGCGACGTTG	TCCCTCCGCA	TAGAGAGTAT	CAAAAGCCAG	4560
ACTGGACTTC	CCTGAACCTG	ACAAGCCAGT	CACGACAACC	AACCTGTCTC	GCGGAATCTC	4620
CACATCAATA	TTTTTTAAAT	TATGGGCACG	CGCCCCATGA	ATGACAATTT	TATCTTGCAT	4680
CTTTGTTCTT	TCTAGTCCAT	TATTGCTTAC	CATTATACCA	AAAAAAGTGA	GATTCTATTA	4740
CCCCAAAGGC	CGATTTTGTA	GTATAATAGT	ACAGTGTGAA	AAAATCTGAA	AAATGAGAAA	4800
GGATAAGGGA	TATGAAACAA	GTTTTTCTCT	CTACAACAAC	TGAATTTAAA	GAGATCGATA	4860
CGCTTGAACC	GGGTACTTGG	ATCAATCTCG	TCAATCCGAC	TCAAAATGAA	TCACTCGAAA	4920



523

TCGCCAACAC	CTTCGATATT	GATATTGCTG	ACCTTCGAGC	ACCGCTCGAT	GCGGAAGAAA	4980
TGTCTCGTAT	TACCATTTGAA	GACGAGTATA	CCCTGATTAT	CGTAGACGTG	CCGGTCACGG	5040
AGGAAAGAAA	TAACCGCACC	TACTACGTAA	CCATCCCGCT	TGGTATTATC	ATCACTGAGG	5100
AAACCATTAT	CACTACGTGT	TTGGAACCAC	TACCTGTCCT	TGATGTCTTT	ATCAACCGTC	5160
GATTGCGTAA	TTTCTATACC	TTTATGCGTT	CACGTTTTAT	CTTTCAAATT	CTTTATCGCA	5220
ATGCAGAGCT	TTACCTAACA	GCCCTTCGTT	CAATCGACCG	CAAGAGTGAA	CAAATCGAAA	5280
GTCAACTGCA	TCAATCAACT	CGTAATGAAG	AATTGATTGA	GCTCATGGAA	TTGGAAAAAA	5340
CTATCGTCTA	TTTCAAGGCC	TCCCTCAAAA	CAAATGAGCG	CGTGATTAAG	AAATTGACCA	5400
GTTCAACCAG	CAATATCAAG	AAATACCTTG	AGGACGAAGA	CCTGCTTGAA	GACACCCTGA	5460
TTGAAACCCA	ACAGGCCATC	GAGATGGCAG	ATATTTATGG	AAACGTCTTG	CATTCTATGA	5520
CAGAGACCTT	TGCCTCTATC	ATTTCTAACA	ACCAGAACAA	CATCATGAAA	ACCTTGCCCC	5580
TTGTGACCAT	CGTCATGTCC	ATCCCAACCA	TGGTCTTTTC	TGCCTACGGG	ATGAACTTTA	5640
AGGATAATGA	AATCCCCCTA	AACGGAGAGC	CAAATGCCTT	CTGGTTAATC	GTCTTTATCG	5700
CCTTTGCTAT	GAGTGTCTCG	CTCACTCTCT	ATCTCATCCA	TAAAAAATGG	TTCTAAGAGG	5760
AGTTCTATG	TCTCAAATTG	ATCTACAAAA	ATCAACTAAG	AAAAACCAAG	AGTTTGTCCTA	5820
CATTGCTACC	CAACAATTCA	TCAAAGATGG	GAAAACAGAC	GCTGAAATCC	AGACTATTTT	5880
TGAGGAAGTC	ATTCCCCAAA	TCCTTGAGGA	GCAATCTAAA	GGTACAACCTG	CCCGTTCCCT	5940
ATACGGCGCA	CCAACCTATT	GGGCTCATAG	CTTCACTGTC	AAAGAGCAGT	ACGAAAAAGA	6000
GCATCCAAAA	GAAAATGATG	ACCCAAAACT	GATGATTATG	GACTCAGCTC	TTTTCATCAC	6060
TAGCCTCTTT	GCCCTTGCTA	GCGCCCTCAC	AACCTTCTTT	GCGGCAGACC	AAGCTTTCGG	6120
CTATGGATTG	ATTACTCTTC	TATTAGTTGG	ACTGGTTGGT	GGATTTGCCT	TCTACTTGAT	6180
GTACTACTTT	GTTTACCAAT	ACTATGGACC	AGATATGGAT	CGCAGTCAAC	GTCCACCTTT	6240
CTGGAAATCT	GTACTAGTTA	TCCTAGCTTC	TATGTTCCCT	TGGTTGCTTG	TCTTCTTTGC	6300
AACAAGCTTC	CTACCAGCTA	GCCTTAACCC	AGTACTGGAT	CCATTGCCAC	TAGCTATTAT	6360
TGGAGCAGCC	CTCCTAGCCC	TTGCTTCTA	TCTCAAGAAA	CGCTTGAATA	TCCGTAGTGC	6420
AAGTGCAGGA	CCAACACGCT	ATCAAGAATA	AGAAAACGAT	AAAAGCAACT	GCAGGTGCGG	6480
TTGCTTTTTT	ACTTACTTTT	TTGAGTTATA	TTCAATGAAA	ATCAAAGAGC	AAACTAGGAA	6540
GCTAGCTGCA	GGTTGCTCAA	AGCACAGCTT	TGAGGTTGCA	GATAAACTG	ACGTGGTTTG	6600
AAGAGATTTT	CGAAGAGTAT	TAAAAGTATT	CTTCTGAAAT	CCCACATAGC	TTTCTCTTAT	6660

524

ATTTTGTGAT AAAATAGGCT CAATCTATTT CTAGGAGGAT GAGATATGGT TTCTACTATT	6720
GGTATTGTTA GTTTATCTAG TGGCATTATC GGAGAGGATT TTGTCAAACA CGAAGTGGAC	6780
TTGGGTATCC AACGTCTCAA GGATCTGGGA CTCAATCCCA TCTTTTT	6827

(2) INFORMATION FOR SEQ ID NO: 61:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11864 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

CTGGCTAGTT GCATAGAGCA AAGTTGCTTC TTCATCAACA AAACCGTTCA TTTCAAATA	60
GGAAAGCAGC TCATCAGGAC TCTCCAAACG AATCCCTTG TAATCCAGCT CAACTGCCAC	120
CTCTTTCAAG GCTGCAAGAA GAAGTGTTC CAGGCCCTGT CTCTGATGGT CAAACTCGAT	180
GACTAAAGAA TGTACTTTTA GACATTGCGG ATTGTCTGAC TGGGGACTTG ATAAATATA	240
GCCTAAAAGT TGATTTTCAT CCCTAGCTAG AAGAAAGGTA TCCGCACACT TACGGATACT	300
TTCTTCTAAA ATATGGGAAA GTTGCTGCTT TTCAGCTGGA AAAGACGAGG TCTGAAGTGC	360
CCCTATCTCA GGCAATCAG ACTTGCTTGC CTGAATGATC TTAATTGGAA TTTCCATGGG	420
AACATCCTAT TGAACATTGC TTGTCAAGTT AGACAAGAGA CGCTCAAATG AGTATTCATA	480
GGTTTGGATG TCTCCTGCTC CCATAAGAC GTAAACAGCA TTGTCATGGT CTAGGAGTGG	540
AGAAACATTT TCAACAGTAA TCACTTGGTG TTTTGTGTTG ATTTGTGTTG CTAGGTCTTC	600
TACCTTAACG TCACCATGAT CTACTTCACG AGCCGAGCCA TAAATTTGCG CTAGATAAAC	660
AGCATCTGCT TGGTTTAAAG CATCGGCAAA GTCCTCCAAC AAGGCAATGG TTCTTGTAAG	720
GGTATGCGGT TGAAAGACTG CTACAATTTC CTGCTTGGG TATTTCTGAC GAGCCGCATC	780
CAAGGTCCGA ATAATTTCTG TTGGATGGTG GGCAAAGTCA TCGATAATCA CTGTATCATT	840
GACAATTTTC TCAGTGAAC GACGTTTAAAC ACCCGCAAAT GTTTTCAAGT GCTCACGCAC	900
CAAGTTCAAA TCAAATCCTG CTGTGTAAAG AAGACCAATA ACGGCTGCTG CATTATGAT	960
ATTGTGACGA CCAAAGGTTG GAATGTGGAA TTGCCCCAAG TTTTGTCCAC GGAAATGAAC	1020
GGTGAAGGTT GAACCAAGTTA TTGAACGAAG AAGATCACTA GCTACAAAGT CATTGCCTTC	1080
AGCTTCAAAA CCATAATAAT AAATTGGTGC ATCAGACGTA ATCTTACGCA ATTCAGCATC	1140
TTCAACCATAG AAAAAAGAC CCTTGGTGAT TTGTTGGCA TAGTCGTAA AGGCATTAAA	1200
AACATCCTCG AGACTTGTGA AATAATCTGG ATGGTCAAAG TCAATGTTGG TGATAATAGA	1260

525

GTATTCTGGG TGGTAAGGCA TGAAGTGACG CTCATATTCG TCAGATTCAA AGACAAAATA	1320
TTTGGCATTG GCCGAACCAC GACCTGTCCC ATCTCCAATC AAGAAGCTGG TATCTGTAAT	1380
GTGAGACAAG ACATGAGACA ACATACCTGT CGTTGAAGTT TTTCCATGTG CTCCTGCTAC	1440
TCCCATGCTA ACAAAGTCAC GCATAAAGCT ACCTAGAAAC TCATGGTAAC GTTTGTAGCT	1500
GATACCATTT TGGTCCGCAT AGGCAATTC GACGTTGTTA TCTGGACGAA AGGCATTTCC	1560
AGCGATAATT TCCATATCAC CGTCTAGATT TTTTTCATCA AAAGGAAGAA TGGTAATTCC	1620
TGCCTGCTCA AGACCGCGTT GGGTAAAGTA GTACTTTTCA ACATCTGATC CCTGAACCTT	1680
GTGCCCCATC TGGTGCAACA TCAAGGCCAA GGCACATC CTGATCCCT TAATTCCGAT	1740
AAAATGATAT GTCTTTGACA TGTTTCTCC CCTATTCTGT CATTCTGGTC AGATTCAAAT	1800
CTTGGGCAAC CCGACGTTCT TGTTCTGTTT GTTTACTTTT TTTATTGTAG ATTTGGCTCT	1860
TCTTTAGAAA ATCATAATTG TTTTCTTTG GAGCAGGTGC TGACACTTCT TCATTCTTGG	1920
TAGGGATAGA ATGAACTTCT TCCGCCAAGA TATAATGAGA CTGGGTCAAT TTTTGGCTAT	1980
ATTTGACAAA TTCACCAGGA TTTTCTTTT GGAAAGGAGC TGTCGGTTGA TTGCCCTGTC	2040
TAAGTAGACT GGGCTGAGAA TGACGTCTCG CAAGGCTGAA ATCCTGAGTT AGGTAGTTAG	2100
CAGAGCGTTT CTTTTCACAG TCCGCACGCG CTTCTTCACG CGCCACCTCC GCATAGCTCT	2160
TTCTTCTTTT TTTAACCCCT AAAGGAGCCT TTTTAGGTTT TTCGACTTGC TTTTCAATCG	2220
GTTTTACTGG TTTTCTTCA GCAATAGGAG CCCATTCTAA ATAATTTTAA TCTCGATACT	2280
CACCCTTGAT ATTACTGATC AGATCAGACT CATCATAGAG ATTCATGACT GGCATTTACAG	2340
TCAACATGAC CTCGTATCT GACACCAATG GAAATCGTTC TTGTTTCATT TTCTATTCC	2400
TTTCAACACT TCATTATAGC GTATTGTCTT GATTTTTCAA GTGCTGGCTT CAGAAATTC	2460
CAAAATTTCT CTAATTTCTG CTAGGGTCAG ACTACCACGT GACTCTGTGC CGTCCAATAC	2520
TTGTGACACC AGATGTTTCT TTTGTTCTTG GAGTTCCTGA ATTTTCTT CAATGGTTCC	2580
CTTGGTCACC AAGCGATAGA CCTCAACCGT TTCTTCCTGA CCCATCCGAT GGGCACGGCC	2640
AATGGCTTGC GCTTCCACCG CAGGATTCCA CCAAAGGTCA ACCAAGATCA CTGTATCTGC	2700
ACCTGTCAGG TTCAGACCGA CCCCACCAGC CTTGAGGGAA ATCAGAAAGG CATCTCTTTC	2760
TCCTTGTTA AAGGCCTTGG TCATGTCTTG TCTTTCCTTG GCTGGGGTTG AACCCGTAAT	2820
TTTAAAGGAA GTCAGGCCCA AGTCTGGCAG TTCTTGTTCA ATTTTTCCTA ACATTCCCTT	2880
GAAGTGAAG AAAATCAAGA CACGGTGTCC GCCGTCTGCC ACCTGTACCA GTAGGTCTCG	2940
GAGACTATCT AGTTTGCCGC TGGCTCCCTG ATAATCTTCC ATAAACAGGG CAGGAGTGTC	3000

526

ACATATTTGA	CGCAAGCGCA	TCAAACCAGA	TAAAATTTCC	ACACGACTTC	GCTGAAATTC	3060
CTGTTCTGAC	ACTTGAGCCA	GATGGTCTCG	CATCTGTTGT	AACTGGGCAA	GGTAAATAGC	3120
CTTTTGCTGG	TCTTCCAGTT	CATTTTATA	AACCACCTCA	ATCAAGTCTG	GCAATTCAGT	3180
CAGAACTTCT	TCTTTCTTGC	GTCGCATCAC	GAAAGGCTTG	ATAAACTGAG	CCACTCGCTC	3240
TGCTGGCAAT	TTCATAAAT	CTTTCTTGCT	TGGCAAAAGT	CCAGGCATGA	CGATTTGGAA	3300
AATAGACCAC	AACTCACCCA	GATGGTTTTC	AATCGGAGTT	CCTGACAAGG	CAAAGACCGA	3360
CGGCACCACA	AATTGTCTCA	AGGTCTGGGC	AATCTTGGTC	TGGGCATTTT	TCATGACCTG	3420
AGCCTCATCT	AAGAAAAGGA	AGTCAAAGGC	CATCCCTTGA	TAAAACTCAC	TGTCCTGACG	3480
GAAGGTGGCA	TAGCTAGTCA	CATAGATTG	ATGGCTCTCG	GCAAGAATCT	CCTCACGACT	3540
TGCTTTCAAA	CCATGAACAA	CAGTCACATC	CAACTGTGGA	GCAAATTTCT	GAAACTCATC	3600
TGCCCAGTTG	TAAATCAAAC	CCGACGGAGC	GAGAATCAAA	ACCCGACTTT	CTTTTGTAC	3660
TTGACTAGTC	AAAAAAGCAA	TGGTCTGAAG	GGTTTTCCCA	AGTCCCATAT	CATCAGCCAA	3720
AATCCCACCA	AAACCATAAT	GATGGAGCAT	CTGCAACCAG	CCAAATCCCT	TTTCCTGATA	3780
ATCTCGCAAG	TCAGCCTTGA	CCTGAGTTGC	TTGCAAAGGA	AAGTCCTCTG	GATGCGTCAA	3840
ATCCTGGGCC	AGATTCTGGA	ATTCTTGTGA	AAAAGAAACA	CGGTCTCGCC	CTTCAAAGAG	3900
ATGAGCTAAA	CTGTAGGCCA	AGGATTTCCG	AGCCTGCAAG	GTCCCACCTT	TTAATTCAAA	3960
TTGCCCCAGT	TCCTGTAGAT	TTTGGCGAAT	TTTCTTGGTT	TCTTCATCGA	AAAAGTAAAC	4020
TTGATTAGAC	GAATCAATAT	AAAAATCCTG	ATTGGCAACC	AAGGCCTGCA	TGGCTTGGTC	4080
GATTTCTCTC	TGGACAATAT	TTTGAAAATC	AAACTGGATT	TCCAAGAGAC	CTCCCTTGGA	4140
GGCAATCTGC	ACCTGAGGAC	TCGCTAGGCT	ATAAAGCTCT	TCTAGTTTAT	CTGATAGGTC	4200
AACATGCCCG	AGTTTTTCAA	AGACTGGAAT	GATATCATGA	AAAAAATGAT	AGACAGACTC	4260
CGCTTTTAAG	GCCTGACGCC	AAGATTGAAA	ATCGGCCTCA	AAGCCCGCAG	CCAAACAGAC	4320
TTGGAAAATT	CTTTCTTCTA	AGTCTGCGTC	ACTTGAAAAG	GGTAATTCTT	CTAGCTCTTG	4380
TCGGCTAGAT	ACCTGTCTAT	TTCCATAATC	AAACTGAATT	TCTAAACGAA	TCCGATTATC	4440
TTCTTCCCTG	TCAAAGTAAA	AAGAGGGCGC	AAAAGTTTGT	ATTTGTAGAC	GTTCTGGAGC	4500
TGAAACGGTG	CCCATCTGGA	TAAAAAGAGT	CAGACAGGAG	GCCAATTTGT	CTCGATCACT	4560
GCTATCAAAT	TGCAGGTATT	TCTTTCCTTG	TTGACCCACA	GGTAACGCTT	TAATTTCCCT	4620
GAGAAGACGC	ATCTGCTGGT	CTGTAAAAAA	ATAAACCTGA	CCTTTATGGA	AAAGTACTGC	4680
TCCCTGATAA	AAGACATTGA	CCCTAGGACT	CTCACTGATT	TCCATTTCAA	AATAATCCGA	4740
GTATTCTGTT	ACTGTAAAGG	CAAATAGATT	GGCATCAGCA	TGCATATCCT	GAAAAAGCAG	4800

527

GGTTTGGTAG CTATCCACTT GATGGTCAAA TTGAAAATGG GGCAAGGCCA TCAGTAAATT	4860
CACACCCTGC TCAAAAAAGG TCAGAGGGAA AAAGAGGTGC CGACCTTGGT TTTGGAAAAA	4920
GAGGTCTGGA ACCAGCCCTT CCTCCGTTAG TCCGTGCAAG AAAGTCAAAA GTTCTTGGCT	4980
GGCATCATCA AAGGCTTCCC AAGAAAGAGA CTCCTCATAA ATCTTGCCAA TCATATACGA	5040
CTTCTCTGTC TCGACAATCC TTAaaaaaaag TGGAATATCC CGAATGACAT AGTATTTTGT	5100
GCTATTGATT TGGCCGATTC TCAGAGTCCA CAAGATATGA TTGGTTCCTG CTTCCACCTG	5160
ACCCACAGCT GATAACTCAT AGGCGCATTC TGATTTTGGG GATAAAATTC GATCCAAAAA	5220
CTTGCCACCC AAGGTCACCT TGGTTTCAAC AGCCTCTTTT TCTTCATGAC CTTCTTCCAG	5280
ACTCCACAAG ATTTCTTGAC CACGCTCATC ATTTTTCAGA AAATGCTCTA GCGCTGCCAA	5340
ATGCACACAG TAGCCCTCTT TTTGAAAAAA ATCAGAGGCA CAAAAACCA AATCATCCTC	5400
TAAACTATAG CGCAGTTCTT CTTCTGCAAC GCGAGCGTAG AGCCGATTGT TCTTTTCCTT	5460
GATGATATCA ACCTTACCAG TTTTATAAAG GGCAACACCT TCGATACGAA TTTTCCCCGG	5520
AATCAATTTA GCCATATTTT CACCTTTACC TTATCTTTTT ATTATACCAT ATTTTCGCCT	5580
ATGAAAATAG CCTTCTAGGA AGACTTTTCT CCTAGAAGGC TGGATTTTGA ACGTTTGGCA	5640
AAAGTAGCCA CAATCCGCTG ACAGACTTCT TGCAACAGAG ATTTGGGCAT AGCTATATTG	5700
ATGCGGGCAT GGAGACTTCC TTCCTCTCCA AAATCCAAAC CACGGTTGAG GATAACCTTG	5760
GCTTCATTTT TCAACAACTC TTGCAATGTT TCATCAGTCA GGTGATAAGC TGAAAAGTCA	5820
AGCCAAATCA AGTAGGTACC TTGCGGTTTC ATGACCTTGA TTTTACTCTC TTTTCCAAAT	5880
AGATCCATCA CATAATTGAT GTGGTCTTCA AAGACTTGCT TGAGTTCCTC TAGCCAATCT	5940
TTACCGTATC GATAGGCAGC TTCTGTGCGC AAATAACCCA AGCCTGAAAT TTTCTGCTGA	6000
TTATTGGCCA ACAGGCGTTT CTGGAAGGCC AGTCTCAACT TAGGATTTTC AATGACTGCA	6060
TAGGAATTTT TTGTTCCAGC AATATTAAAT GTTTTAGTGG CACTGCTCAA GACGATAGCA	6120
AAATTTTGA AGGCAGGATT GATGGTATTG AAAGACTGGT GTTTGTGACC AAAGAGGCTC	6180
AAATCTTGGT GAATCTCATC CGAAACTAAC AAAACACCGT CTTTTTGGCA GAGTTGGCCA	6240
ATCTTCTCCA ACACTTCTTT TTCCCAAACA CGTCCACCAG GATTGTGAGG GTTGCAAAGA	6300
ACATAGAGTT TAACCTCCTC TTCCACCAAA TCCTTTTCAA GTTGGTCAAA GTCAATCTCA	6360
AACAGACTAT CCTTTTCCAC TAAGGAATTA GTAATCAATC TACGATTATT CAACTTGACA	6420
CTGCGAGCAA AGGGTGGGTA GACAGGCGTG TTAATTAAAA CCGCCTCGCC TTCTTTTGTA	6480
AAGGTTTGAA TAGCTGTTGA GATGGCTGGT ACCACACCCT CGATAAAGAC AAGAGCCTCT	6540

528

TTGTCAAAGT TGTAACCGTA TTGTGTAGCT TCCCACTTTT GAACTTCCTT AATTAAGTCT	6600
TCACTGGCAT AGGTATAACC ATAAACCAGT TGGTCTGCGT AAGTTTGCAC GGCTTGGCGG	6660
ATTTTCAGGCA AGACCACAAA GTCCATATCC GCTATCCAAG CTGGTAGAAC TTCACTATCC	6720
GTTTCTGTTT CTTTCCATTT ATAGGTATGG TGCCCTAAAC GGTTGGGCAG GCTTGTAATA	6780
TCATATTTTC CCATCTTTGT CTTATCCTTC TATGGCTTGG CGCAAATCTG CAATCAAATC	6840
TCTAGCATCC TCAATCCCAA TAGACAAACG CAAGAGGTCA TCTGTCAAAC CATAAGAATG	6900
GCGTACCTCT GCTGGAATAT CAGCATGAGT TTGAGTCGTT GGATAAGTAA TAAGACTTTT	6960
CACCTCCACCC AAACCTTTCCG CAAAAGAGAA GACCTTGAGA CTGTTCAAAA TATGAGGAAT	7020
GCGTGTTTCA TCGGCTACTT TAAAGGAAAT CATGCCTCCA CGACCAGTGT AGAGAACTTC	7080
CTTAACTGCT GGAGAATCCT TCAAAAAGGC AACCACCTCT TGGGCGTTAG CTGTTGAGCG	7140
CTCCATACGA AGAGACAAGG TCTTGAGACC ACGAAGCAAC TGGTAGCTGT CAAATGGAGA	7200
CAAGACTGCC CCTGTTGTAT TAAGATTGTA AAAAAGCTTC TCGTATAGTT CTAAACTATT	7260
GGTCACAACC ACTCCAGCCA AGACATCATT GTGGCCTGCT AGATACTTGG TTGCTGAATG	7320
GAGAACGATA TCTGCTCCAT CTTCAATCGG ACOTTTGGTAG ATAGGGCTAT AGAAGGTATT	7380
GTCCACCACC ACTTTGGCAC CCTTAGCATG AGCCAATTTT GCTAGTTTTT CGATATCAAA	7440
TTCCAACATC AAGGGATTGG TTGGGGTTTC GATATAGAGA ACATCCACAT CTTTTTCTAA	7500
CTCGGCAATC AACTCTTCTT CTGTATTGGC ATAGGTAAAA TGGAAATGAC CTTCTGCTC	7560
CACCTGGTTA AACCAGCGAA AAGAACCACC GTAAAGATCA CGCACTGCCA AGACCTTACT	7620
TCCTACTGGA AAGACGCTAA AGGCCAGTAC AATAGCTGAC ATCCCTGAGC TAGTCGCTAG	7680
GGCATAGCTT GCTGACTCAA TAGCCGCCAA GACTTCCTCA GCCTTACTAC GAGTTGGATT	7740
TTTAGTGCGC GTATAGTCAA ACCCAGTAGA TCGACCAAAC TCTGGATGCT CATAGGTGCT	7800
TGAAAAATGA AGTGGTGTCA CCAAAGCACC TGTTCCTCA TCAGACTTGA TCCCTGCTTG	7860
TGCTAAAATT GTGTTAATGT GTAATTCCTT GCTCATACAA TTCTTCCAAA TCTATAGTAA	7920
CTATTGTACC ACTTATTTTG TATCCTTCGT TTTCTTGTTC TCAAGAGCTA GTTATAGTTT	7980
CAAACTATAT AAAAAGGGAG TTTTTCCTGC TCCCTTTAAT AGACTATAAA ATGGTGAATC	8040
TCAAAAGACA CCTTCACTCT ATCATTTGCT CCTGCACAAA ACGAGCATAA CGCTCATGAT	8100
TTTCCAGTAG TTCCTTATGA GTTCCTGAGC CAGTGATTTT CCCCTCCTCT AAGAAGAAAA	8160
TACAATCCAC ATCTTTTACC GTTGACAAAC GATGCGCTAT AATCACAACC GTCTTCTCCT	8220
TTAGTACAGA ATAGAGGCTA CTGATAATCG CATACTCAGA ATCCGCATCA AGATTAGCAG	8280
TGGCTTCATC AAATATAAGA ATTCAGCAT CTTTAAGTA GGCTCTAGCT ATTTGAAGTC	8340

TTTCGTTCGC CCCCTGACA AGAGTCGTCC GCGTTCACCA ACTTCAGTAT CTAGTCCCTC	8400
TTTCATGGAG CGAATCTCAT CACCTAGTGA TACTAAGTCT AGCACTTTCA TCAATTCATC	8460
ATCAGTTACT AAGCGATTCA AACCGAGACA AAGATTGTCA CGAATACTGC CAGATAAGAC	8520
TGCATTATTT TGTGAAACCC AAGCGATTTT ACTTCTCCAT TCTTTTAAGT TAAAATCATA	8580
TATACTTGAT TGCTCCATTA GAATATCTCC TGAAAGCGGT TTATAAAACC GCTCTAACAA	8640
ACGCACAATC GTTGATTTTC CTGATCCAGA TGGTCCAACA AAAGCAATTT TTTGCCCTT	8700
GAAAATTGAA CAAGTAATAT CCTTTAAGAC AGGTGCGATT TCATCATAAC CAAAATAGAC	8760
ATGGTTAAAA TTCAACCTTC GTCCTGATAC CGATTTTCCT CCCTCAAATT TTTCTTAGG	8820
AACTGCAAGC AAGTTCTCCA GTGCAACTGA AGATCCCTTG CTCCTAGAAT AAACAGTTAC	8880
AAAATTAGCT ATATTACTAA TAGGATTAAG TAATTGAAAG AGGTAAATCA AAAACGAAAC	8940
CAAGGTTCCC ACAGATATAT ATCCTGCGCT GACCCGATAA CCCCATAGG TTAGCATCAC	9000
AGCTATAGTC GCAAAGATAA ATAAGAGAGC AAACGGGGTC TCAAAAGAAG TAACCTATC	9060
TGATTTAGT GAATTGTTTT GTACCTTTT AATACAATTA TCCAAAACAT CCTGTACACT	9120
TTTCTCTGCT TGGTTAGTCT TAATTAATTC ATGTTCTTGA ATCTTTTCAG TCAATTGCCC	9180
TGTTAAATTT CCTCCTGTAA ACGACGACTA TACTTTTCAC TGATATTGGA AAGGGGCAAG	9240
ATAATAAACA TCATACAAGG AAGAGTGATG AATAAAAGTA GAGAAAGATT CCAATCAAGA	9300
CTAAATAAGA CTACAATGGA ACCAAGTACC ATAACATAAC TCAGAATAAT ATTTGGGAAA	9360
GTCGTAATTA AAAACTCAGG AATGACACTC GTGTCATTGA CAATGGCAGA AGTCAACTCC	9420
CCACTTTGGC TCTTATCAAA GAAGGATTTT TCTACATAAA TCAACCCCTC TATCACTTTT	9480
TTCCGTGATT TTGCTATCTT TTTTTCACCC GATTGACTAA ACAGATAGTA ACCAATAGAA	9540
GAAAACAAGG CTTGACCAAT AAAAATCAAA AACGATTGAA ATACTTTGGA GCCTATATTT	9600
TCAATAGAAC TCCCATCTAT TAAATCCTTT AAGATAAGGG GAAGCAACAA AGCAAGTAGA	9660
CTAGACAGAA CAAGTAAGAA ACTCCCCATA ATCACCTTAG TATCTACTCT TAATAATTTT	9720
AATTTCATAA ATACTCCTTA TAATATTTCA ACGGATAAAG TCGGGAATAA CTCAATTTGA	9780
GGATAAAATC TAATAAATCT TCCTATAACA AAACGCATAA CATCTAGGAT TTTATATACC	9840
TGATATTATG CGTTTTTAAG CACAAAGACT TCTTACACAA ACTTATCTAC AATTAGATTT	9900
TATTTGACAT GTTTTGCCAA TTCTTCTTGG GCTTTTTTAT TGGATTCTTC TTTTCTTTC	9960
AACCATTTT CTCTGGCTTT TGCATATTCG TCTGTTGTGA CAATCTTATC TTGTACTTTG	10020
AGGTATTTAT ATGATTCAAC CCCTTTTGTA CCGGTTAAAC CATAGGCAGC AGCAAATGGT	10080

530

ACGGTTCTTC	TCAATGATGG	TGTTCCCCCA	CGCGAAACAC	TTGGAAGAAC	TAAAGAACTA	10140
TCAATCAACC	AAGCTTGAAT	ATCAGCATAT	TTCTCATAAC	GTTTGGCCGG	ATCTTGCTCT	10200
TTATTAGCTT	CTTCCAACAT	TTGAGTATAG	ACATCCAGTC	CAACTGCCTT	AGCCTTGTC	10260
TTGGCCTCAC	CAGGCTCTAG	TCCAAGATTT	TGCAGAAATC	CTCCACTATT	AGTATTAAAA	10320
ATATCGAGAT	AGGTTGACGG	GTCTTGATAA	TCAGGTCCCC	AACCGCCATG	ATATAAATCA	10380
TAATCTTTCT	GAGCAGCTGT	TTGAGCAAAG	TAGCCTGAAC	TGTCAAACTC	ATCTGATGTT	10440
AATTGCTGAA	TGTCAATCAC	TACATTATCA	GAACCTAAAA	CAGATTCAAT	TGATTGTTTG	10500
ATAGAACTAA	CTCCTTGAT	GCCTACTTTA	TCTGTTACTT	CCACAGCTCT	ATCCAAGTGG	10560
ATTGGGAATT	GAACACCCTT	TGCTTCGAGT	TCTTTCTTAG	CTTCGCAAAA	CTTAGCCTTG	10620
GCTTTCTCAG	GATTGTAGTA	AGGGTCTTGA	CCATCCGCAA	AGTTGATACC	TTGCCATTCC	10680
TTACCATAGT	TGACCATCTT	AGAGGCTACA	ACTTCACCAA	AGTCTTTTCC	CTTGATACTG	10740
ACAAAGTTTG	GAGGAACCAC	TAGGTTACGC	AAAATCTTTG	TTGCACCTTC	TTTCCCTTCA	10800
GACTGAGCCC	CATAAGATGT	TCTGTCAAAA	GCAAAATTGA	TAGCCTGACG	GAAGTTTFTA	10860
TTGAGAACTG	CTTCCTGAGT	CGATTTCTTT	TCAATGTCAC	TTGTTTLAGA	AGTATAATTG	10920
TAAGACTTCC	TATCTAGGTT	AAAATTAAAG	AAATATGAAG	TTGAATTTTG	CATACTATAG	10980
ATGATATTGT	TTTTGTATTT	TTCTTTAATC	CCTTCATAGC	TGGAGCTGTT	AGGAAAAAGA	11040
CGAGCCGTAG	TATAAGCACC	AGCTGTAAAA	TTACGTTCCA	GTGATTCTTG	GTCSCTACCA	11100
TCATAGTAGG	TCAATTTTCA	ATCGTCTACA	AAGACATTCT	TAGCATCCCA	GTAATTAGGG	11160
TTTTTCTTAT	ATTCAATAGC	AGATTTTGAG	ACAAGTGCTT	TCATCAAGAA	AGGTCCATTG	11220
TACAAAATAC	TAGATGGATC	CGCCTTCCCA	AAATCATCCC	CTTTTGATTT	CAGGAAATCT	11280
GCATTAAACAG	GAAAAAGTAT	CGTTGCAAGT	GTTTTTGAAT	TCCAGTAAAG	TTCTGGTTTA	11340
ACCAAAGTAT	ATTGAACCGT	TTGGTCATCA	AGTGCCTTGA	CACCGACAGT	TGAAAAGTCS	11400
CTTGTTTTAC	CAGTGATATA	GTCATCCAAA	CCAGCAACAG	AGTCCTGCAC	TAGATACAAG	11460
GCTTCTGATT	TTTTATCAGC	TGCATATTGC	AAACCTGTCA	CAAAATCCTG	GGCAGTTACA	11520
GGCGCATATT	CTTCTCCCTC	AGAAGTAAAC	CACCTGGCAT	CCTTACGAAG	TTTGTAGGTA	11580
TAGGTCAAAC	CGTCTGAGA	AACAGTCCAA	TCCTCTGCTA	ATGATGGAAT	AATATTCCCA	11640
TATTGGTCAT	TTTCTAATAA	CCCGTCTACC	AAATTTGCAA	CAATATCGGA	TGTTGCTGCG	11700
CGGTTTTCTG	CTAGATAGTT	CAAGCTAGAT	GGATCACTTG	AATAAACATA	GTTGTAGGTT	11760
TTTGACGCCG	TGCTAGAATT	TCCACACGCG	CTCAATAAAA	CTCCTGTACC	CAGGACAAGA	11820
CCTGCCAAGG	TTAGATATTT	GCTCTTAGAC	TTTTTCAATT	CCGG		11864



531

## (2) INFORMATION FOR SEQ ID NO: 62:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2412 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

TAAGTGCCTT AACATAATA TAAGGAGAGA AAATGCTCTGC AATAGAACGT ATTACAAAAG	60
CTGCTCACTT AATTGATATG AACGATATTA TCCGTGAAGG GAATCCTACT CTACGCGCGA	120
TTGCTGAGGA AGTCACTTTC CCCCTATCTG ACCAGGAAAT CATCCTAGGC GAAAAGATGA	180
TGCAATTCCT TAAACATTCC CAAGATCCTG TCATGGCTGA AAAAATGGGA CTCCGCGGTG	240
GTGTTGGACT GGCTGCTCCC CAGTTAGATA TCTCAAAACG CATTATCGCT GTTTTGGTAC	300
CTAATATTGT TGAAGAAGGC GAAACTCCAC AGGAAGCCTA CGATTTCGAA GCCATTATGT	360
ACAATCCAAA AATCGTCTCT CACTCTGTTC AAGATGCTGC TCTTGGCGAA GGAGAAGGTT	420
GCCTGTCTGT TGACCGTAAC GTGCCCTGGCT ATGTTGTTCC CCATGCCCCG GTTACTGTTG	480
ACTACTTTGA CAAAGATGGA GAAAAACACC GTATCAAAC CAAAGGCTAC AACTCCATTG	540
TTGTTTACGA TGAAATTGAC CACATTAACG GTATCATGTT TTACGATCGC ATCAATGAAA	600
AAGACCCATT TGCAGTTAAA GATGGTTTAC TGATTCTTGA ATAAAGAAAA TCCCGTTGCA	660
AGACGGGGTT TTGTGTTATA ATAGAGGCAT GAAAAACAAAT GATATTGTCT ATGGTGTCCA	720
CGCCGTTACC GAAGCCCTCC TTGCAAATAC AGGAAACAAA CTCTACCTCC AAGAAGATCT	780
CCGAGGTAAG AATGTTGAGA AAGTCAAGGA ACTAGCTACA GAAAAGAAGG TGTCCATTTC	840
TTGGACATCA AAAAAATCTC TCTCTGAGAT TACTGAAGGT GCTGTTTCATC AAGGTTTTGT	900
TCTACGAGTG TCTGAATTTG CCTATAGCGA GCTAGATTAC ATCCTTGCAA AAACACGCCA	960
AGAAGAAAAT CCACTTCTAT TGATTCTAGA TGGTCTAACC GATCCCCATA ATCTGGGTTT	1020
TATCTTGCGA ACAGCCGATG CGACCAATGT TTCAGGTGTC ATCATTCCCA AGCACCCTAC	1080
TGTCGGAGTA ACTCCTGTCG TTGCCAAAAC AGCCACAGGT GCTATTGAAC ACGTCCCAAT	1140
TGCCCCGAGT ACCAACCTCA GTCAAACCTT AGGATAAACT TAAGGATGAA GGTTCCTGGA	1200
CCTTTGGAAC GGATATGAAC GGTACTCCTT GCCACAAGTG GAATACAAAA GGGAAAATCG	1260
CCCTCATCAT TGGAATGAA GAAAAGGTA TCTCTAGCAA CATCAAAAAA CAGGTCGATG	1320
AAATGATTAC CATTCCGATG AATGGACATG TTCAAAGCCT TAATGCCAGT GTTGCTGCGG	1380

532

CCATTCTCAT GTACGAAGTT TTCCGAAATA GACTATAAAA AAGTTTCCAG TCATCTGATT	1440
GGAAACTTTT TTATGATTAA CTATGTTCTG TAATGAATTT ATAGGCTTCT TGACCAGCGA	1500
TAGCTCCATC TCCAACCGCT GTTGTTACTT GCGAAGGTC TTTCAAGCGA ACATCTCCAA	1560
CTGCAAAGAT ACCGTCGACT GCAGTTTCA TGTGGTTATC TGTCACAATC CATCCTGCCT	1620
GATCTTGGAT ATTCAATTCT TTAACAAAAT CGCTAAGAGG GTCCAAACCA ACATAGATAA	1680
AGACACCACC GAAGGCTTGT TCTGTCACTT GACCTGTTTT CACATTTTCA AATACGACTG	1740
ATTCTACTCG GTTTTCACCC TTGATTTCCC TTAACACAGA ATCCCAGATA AAGCTGATTT	1800
TTTCATTTCG AAAGGCGCGA TCTTGTAATA CCTTTTGGGC ACGAAGTTGG TCACGACGGT	1860
GAACAATGGT AACAGTCTTA GCAAAACGAG TCAAGAAGAG GGCTTCTTCA ACAGCTGAAT	1920
CTCCACCACC AACTACCAAT AAATCTTGGT CACGGAAGAA AGCACCATCA CACACAGCAC	1980
AGTAAGAAAC ACCACGACTG TTCAGTTCTT CTTCTCCAGG CACTCCCAA GGACGGTGTT	2040
TAGAACCAGT TGCTACGATA ACTGTACGTG TTTCATATGT TTGGTCATCA GTCATCACTT	2100
TCTTAAATC ACCATGGCTT CGACATTTT AACATAACCA TAAATGTGCT CAACACCAAG	2160
ATTTTCAAGT GGTTCAAACA TCTTTTCAGC CAATTCAGGT CCACTAATAT TAGCGTATCC	2220
TGGGTAATTT TCGATATCAG ATGTATTATT CATCTGACCA CCTGGCAGAC CACCTTCAAT	2280
CAAAGCTACT TTTAGATTGC TTCCAGCAGC ATACAAGGCC GCAGTCATCC CTGCAGGTCC	2340
AGCACCGATA ATAATAGTAT CGTACATATA GATTCCTTCT TTCTTGGTGT AACTATCTTT	2400
ATTCTAACTC TG	2412

(2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7760 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

CCGATTTGGT GGAATTTTGT TCTCATCATT TAGAAGGTGT TGCAAGAGCA GAGTTTACCT	60
TGGTGCTTCA TACCAAATTG GGAGAAGCCT CTGTTTGGC AAATATTGTA GATGTAAACA	120
AGGATGAATG GATTTTAGGA ACAGTTGCTG TGCCAATAC CTTATTGGTT ATTTGTCGAG	180
ATCAGCACGT TGCCAACTC ATGGAAGATC GTTGCTAGA TTTGATGAAA GATAAGTAAG	240
GTCTTGGGAG TTGCTCTCAA GACTTATTTT TGAAAAGGAG AGACAGAAAA TGGCGATAGA	300
AAAGTTATCA CCCGGCATGC AACAGTATGT GGATATTAAA AAGCAATATC CAGATGCTTT	360

533

TTTGCTCTTT CGGATGGGTG ATTTTATGA ATTATTTTAT GAGGATGCCG TCAATGCTGC	420
GCAGATTCTG GAAATTTCCT TAACGAGTCG CAACAAGAAT GCCGACAATC CGATCCCTAT	480
GGCGGGTGTT CCCTATCATT CTGCCCCACA GTATATCGAT GTCTTGATTG AGCAGGGTTA	540
TAAGGTGGCT ATCGCAGAGC AGATGGAAGA TCCTAAACAA <u>GCAGTTGGGG</u> TTGTTAAACG	600
AGAGGTGTGT CAGGTCATTA CGCCAGGGAC AGTGGTCGAT AGCAGTAAGC CGGACAGTCA	660
GAATAATTTT TTGGTTTCCA TAGACCGCGA AGGCAATCAA TTTGGCCTAG CTTATATGGA	720
TTTGGTGACG GGTGACTTTT ATGTGACAGG TCTTTTGGAT TTCACGCTGG TTTGTGGGGA	780
AATCCGTAAC CTCAAGGCTC GAGAAGTGGT GTTGGGTTAT GACTTGTCTG AGGAAGAAGA	840
ACAAATCCTC AGCCGCCAGA TGAATCTGGT ACTCTCTTAT GAAAAAGAAA GCTTTGAAGA	900
CCTTCATTTA TTGGATTTCG GATTGGCAAC GGTGGAGCAA ACGGCATCTA GTAAGCTGCT	960
CCAGTATGTT CATCGGACTC AGATGAGGGA ATTGAACCAC CTCAAACCTG TTATCCGCTA	1020
CGAAATTAAG GATTTCTTGC AGATGGATTA TCCGACCAAG GCTAGTCTGG ATTTGGTTGA	1080
GAATGCTCGC TCAGGTAAGA AACAAGGCAG TCTTTTCTGG CTTTGGGATG AAACCAAAAC	1140
GGCTATGGGG ATGCGTCTCT TCGGTTCTTG GATTCATCGC CCCTTGATTG ATAAGGAACG	1200
AATCGTCCAA CGTCAAGAAG TAGTGCAGGT CTTTCTCGAC CATTTCTTTG AGCGTAGTGA	1260
CTTGACAGAC AGTCTCAAGG GTGTTTATGA CATTGAGCGC TTGGCTAGTC GTGTTTCTTT	1320
TGGCAAAACC AATCCAAAGG ATCTCTTGCA GTTGGCGACT ACCTTGCTA GTGTGCCACC	1380
GATTCGTGCG ATTTTAGAAG GGATGGAGCA ACCTACTCTA GCCTATCTCA TCGCACAAC	1440
GGATGCAATC CCTGAGTTGG AGAGTTTGAT TAGCGCAGCG ATTGCTCCTG AAGCTCCTCA	1500
TGTGATTACA GATGGGGGAA TTATCCGGAC TGGATTTGAT GAGACTTTAG ACAAGTATCG	1560
TTGCGTTCTC AGAGAAGGGA CTAGCTGGAT TGCTGAGATT GAGGCTAAGG AGCGAGAAAA	1620
CTCTGGTATC AGCACGCTCA AGATTGACTA CAATAAAAAG GATGGCTACT ATTTTCATGT	1680
GACCAATTCG CAACTAGGAA ATGTGCCAGC TCACCTTTTC CGCAAGGCGA CGCTGAAAAA	1740
CTCAGAACGC TTTGGAACCG AAGAATTAGC CCGTATCGAG GGAGATATGC TTGAGCGCGG	1800
TGAGAAGTCA GCCAACCTCG AATACGAAAT ATTTATGCGC ATTCGTGAAG AGGTCGGCAA	1860
GTACATCCAG CGTTTACAAG CTCTAGCCCA AGGAATTCGG ACGGTTGATG TCTTACAGAG	1920
TCTGGCGGTT GTGGCTGAAA CCCAGCATTT GATTCGACCT GAGTTTGGTG ACGATTACAA	1980
AATTGATATC CGGAAAGGGC GCCATGCTGT CGTTGAAAAG GTTATGGGGG CTCAGACCTA	2040
TATTCCAAAT ACGATTGAGA TGGCAGAAGA TACCAGTATT CAACTGGTTA CAGGGCCAAA	2100

534

CATGAGTGGG AAGTCTACCT ATATGCGTCA GTTAGCCATG ACGGCGGTTA TGGCCCAGCT	2150
GGGTTCTCTAT GTTCCTGCTG AAAGCGCCCA TTTACCGATT TTTGATGCGA TTTTACCCG	2220
TATCGGAGCA GCAGATGACT TGGTTTCGGG TCAGTCAACC TTTATGGTGG AGATGATGGA	2280
GGCCAATAAT GCCATTTTCG ATGCGACCAA GAACTCTCTC ATTCTCTTTG ATGAATTGGG	2340
ACGTGGAAC T GCAACTTATG ACGGGATGGC TCTTGCTCAG TCCATCATCG AATATATCCA	2400
TGAGCACATC GGAGCTAAGA CCCTCTTTGC GACCCACTAC CATGAGTTGA CTAGTCTGGA	2460
GTCTAGTTTA CAACACTTGG TCAATGTCCA CGTGGCAACT TTGGAGCAGG ATGGGCAGGT	2520
CACCTTCCTT CACAAGATTG AACCGGGACC AGCTGATAAA TCTACGGTAT CCATGTTGCC	2580
AAGATTGCTG GCTTGCCAGC AGACCTTTTA GCAAGGGCGG ATAAGATTTT GACTCAGCTA	2640
GAGAATCAAG GAACAGAGAG TCCTCCTCCC ATGAGACAAA CTAGTGCTGT CACTGAACAG	2700
ATTTCACTCT TTGATAGGGC AGAAGAGCAT CCTATCCTAG CAGAATTAGC TAAACTGGAT	2760
GTGTATAATA TGACACCTAT GCAGGTTATG AATGTCTTAG TAGAGTTAAA ACAGAACTA	2820
TAAAACCAAG ACTCACTAGT TAATCTAGCT GTATCAAGGA GACTTCTTTG ACAATTCTCC	2880
ACTTTTTTGC TAGAATAACA TCACACAAAC AGAATGAAAA GGAGCTGACG CATTGTGCTT	2940
CCCTTTTGTC TATTTTTTAA GGAGAAAGTA TGCTGATTCA GAAAATAAAA ACCTACAAGT	3000
GGCAGGCCCT GGCTTCGCTC CTGATGACAG GCTTGATGGT TGCTAGTTCA CTTCTGCAAC	3060
CGCGTTATCT GCAGGAAGTC TTAGGCGCCC TCCTTACTGG GAAATATGAA GCTATTTATA	3120
GTATCGGGGC TTGGTTGATT GGTGTGGCCG TAGTCGGTCT AGTTGCTGGT GGACTCAATG	3180
TTGTCTCGC AGCCTATATT GCCCAAGGAG TTTCATCCGA CCTTCGGGAG GATGCCCTCC	3240
GTAAAATTCA AACCTTTTCT TATGCTGATA TTGAACAATT TAATGCGGGA AATCTAGTCG	3300
TTGGAATGAC AAATGATATC AACCAGATTC AGAACGTTCT CATGATGACC TTCTAAATTC	3360
TTTTCAGACT TCCCCTCTTG TTCATCGGTT CGTTTATCCT AGCGGTTCAA ACCTTACCTT	3420
CTCTGTGGTG GGTGATTGTT CTCATGGTAG TCTTGATTTT TGGTTTGACT GCTGTCATGA	3480
TGGGAATGAT GGGGCCTCGT TTGCCAAGT TTCAAACCTT TCTTGAGCGC ATCAATGCCA	3540
TTGCCAAGGA AAATTTACGT GCGGTTTCGT TGGTCAAGTC CTTTGTCCAA GAAAAAGAGC	3600
AATTTGCTAA GTTTACAGAG GTCTCAGACG AGCTTCTTGG TCAAAACCTT TACATTGGTT	3660
ATGCCTTTTC AGTAGTGGA CCCTTTATGA TGTGTTGGT TACGGGGCG GTCTTCCTCT	3720
CTATTTGGCT GGTGCGGGA ATGGTTCAGT CGGATCCGTC TGTGTTGGT TCCATCGCTT	3780
CTTTTGTTAA TTACCTAAGC CAGATTATCT TTACCATGT TATGGTTGGA TTTTGGGAA	3840
ATTCTGTCAG CCGTGCCATG ATTTCCATGC GTCGTATTCG AGAAATCTT GACGCAGAGC	3900

535

CAGCTATGAC CTTCAAGGAT ATCCCAGATG AAGAGTTGGT TGGAAGTCTT AGCTTTGAAA	3960
ATGTGACCTT TACCTATCCA ATGGACAAGG AACCGATGCT GAAAGATGTG AGCTTTACTA	4020
TTGAACCTGG TCAAATGGTT GGTGTAGTTG GAGCGACTGG TGCAGGAAAG TCAACCTTGG	4080
CTCAATTGAT TCCACGTCTC TTTGATCCAC AGGACGGGGC CATTAAAATC GGTGGCAAGG	4140
ATATTCGAGA AGTGAGTGAA GGAACCCTGC GTAAAACAGT TTCCATCGTT CTCCAACGTG	4200
CCATTCTTTT TAGTGGAACG ATTGCAGATA ACTTGAGACA GGGGAAGGGG AATGCTACTC	4260
TATTTGAAAT GGAGCGCGCA GCCAATATTG CCCAGGCTAG TGAATTCATT CATCGTATGG	4320
AGAAAACCTT TGAAAGTCCA GTTGAAGAAC GGGGAACCAA TTTCTCTGGT GGACAAAAAC	4380
AAAGGATGTC GATTGCGCGT GGGATTGTCA GCAATCCACG TATTCTGATT TTTGATGATT	4440
CGACCTCAGC CTTGGATGCC AAATCAGAGC GCTTGGTGCA AGAAGCTTTG AATAAGGACT	4500
TGAAGGGGAC GACAACCATT ATTATTGCTC AAAAAATTAG CTCGGTTGTC CATGCAGACA	4560
AGATCTTGGT TCTAAATCAA GGACGATTGA TTGGTCAAGG TACGCATGCA GACTTGGTTG	4620
CCAACAATGC CGTTTACCGT GAAATCTATG AAACACAGAA ATGAAAGACA AACTATAAGA	4680
AAAGTCAATA GTTTTATCTA AACTATTTCT TATTTCAATT TGATGATTTG GCGATGATTT	4740
TAGAGCACGG CAAAAAGCCC TTGAAAAAGT CCATTTTTTC AAAGGTAATC CTGTGTTAAT	4800
TTCAAGAAAT ACATCACTTT TTGTTGCTCA AATGGCAGCT CTTTTTTTAG GATATAAAAC	4860
AGGGTTCCGA TAAGTTTTTT TGCAAGGTGG ATGATGGCTA CATTGTAATG TTTTCCTTGT	4920
TCTAATTTAG TCTTAAGATA GGCCTTAAAA GCAGGCGAAA AGCGAGGGCA TGCTTTGGCA	4980
GCTTGATGA GTACCTACCG CAGATGAGGG GAACTCCGTT TGACCATTCT TCCTGCTAAA	5040
TCAATCTGAT CTGACTGATA AATAGAAGAA TCCAGTCCAG CGAAAGCTTG TAATTGAGCA	5100
GGATTATCAA AGGCATGAAT ATTTGGAATC TCAGCTAAAA TGACCGCCCC TAAACGATCC	5160
CCAATCCCAG TAACCGTCGT GATGACCGAG TTGAACTCAG CCATCAAGTC ATTGACACAT	5220
GTTCGCCCT TGTCAATGAG CCTCTGTAA TGTGTGATGT TTTCAATTACA CGAGATAAAA	5280
CGTCTATGCG TTATCAAACCT CATTACCAAT TAAAACAAAA AGCTGTGGTT AGATCCTTTC	5340
GGAAATTGTC AAGCGATTGG AGGAAATGAA CTAATCCACA GCGGCTTATT CCAAGTATAC	5400
CACTTGGGCT TTGGCAGTAG CTAAGTGGC TAAATATAAT ATAAGGAGGA GTAAAATGAA	5460
GACAGTTCAA TTTTCTTGGC ATTATTTTAA GGTCTACAAG TTCTCATTTG TACTTGTGAT	5520
CCTGATGATT GTTCTGGCGA CTTTGGCCCA AGCCCTCTTT CCAGTCTTTT CTGGACAAGC	5580
GGTGACGCAG CTAGCCAATT TAGTCAAGC TTATCAAAAT GGCAATCCAG AACTTGTATG	5640

536

GCAAAGCCTA TCAGGAATCA TGGTCAATCT TGGCCTGCTG GTTTTGGTTC TATTTATCTC	5700
TACTGTAATA TACATGTGTC TCATGACGCG CGTGATTGCA GAATCGACCA ACCGATGCG	5760
CAAAGGCCTC TTTGGTAAGC TTGCTCAGTT GACGGTTTCT TTCTTTGACC GTCGACAAGA	5820
TGGCGATATC CTGTCTCATT TTACCACTGA TTTGGATAAT ATCCTCCAAG CCTTTAACGA	5880
AAGCTTGATT CAGGTCATGA GCAATATTGT TTTATACATT GGTCTGATTG TTGTCATGTT	5940
TTGAGAGAAAT GTGACGCTGG CTCTCATCAC CATTGCCAGC ACCCCATTGG CTTTCCTTAT	6000
GCTGATTTTC ATCGTGAAAA TGGCACGCAA ATACACCAAC CTCCAGCAGA AAGAGGTAGG	6060
GAAGCTCAAC GCCTATATGG ATGAGAGCAT CTCAGGCCAA AAAGCCGTGA TTGTGCAAGG	6120
AATTCAAGAG GATATGATGG CAGGATTTCT TGAACAAAAT GAGCGCGTGC GCAAGGCAAC	6180
CTTTAAAGGA AGAATGTTCT CAGGAATTCT TTTCCCTGTC ATGAATGGGA TGAGCCTGAT	6240
TAATACAGCC ATCGTCATCT TTGCTGGTTC GGCTGTACTT TTGAATGATA AGTCTATTGA	6300
AACAAGTACA GCCCTAGGTT TGATTGTTAT GTTTGCACAA TTTTCACAGC AGTACTACCA	6360
GCCTATTATC CAAGTTGCAG CGAGTTGGGG AAGCCTTCAG TTGGCCTTTA CTGGAGCTGA	6420
ACGAATTCAG GAAATGTTTG ATGCAGAGGA GGAAATCCGA CCTGAAAAGG CTCCAACCTT	6480
CACTAAGTTG CAAGAAAGTG TTGAAATCAG TCATATCGTT TTTTCATACT TGCCTGATAA	6540
ACCTATTTTG AAAGATGTCG GCATTTCTGC CCCTAAAGGC CAGATGACAG CAGTTGTGCG	6600
GCCGACAGGT TCAGGAAAAA CGACTATTAT GAACCTCATC AATCGCTTTT ATGATGTTGA	6660
TGCTGGTGGT ATTTATTTTG ATGGTAAAGA CATTGCTGGC TATGACTTAG ATAGTCTTAG	6720
AAGCAAGGTG GGAATTGTAT TGCAAGATTC GGTCTTGTTC AGCGGAACGA TTAGAGACAA	6780
TATCCGATTT GGTGTGCCAG ATGCTAGTCA GGAAATGGTT GAGGTAGCAG CAAAAGCAAC	6840
CCACATTCAC GACTATATCG AAAGTTTGCC TGATAAGTAC GATACTCTTA TTGATGATGA	6900
CCAGAGCATC TTTTCAACAG GGCAGAAGCA ATTGATTTCG ATCGCTCGAA CCCTGATGAC	6960
AGATCCAGAA GTTCTCATTG TCGATGAAGC AACTTCAAAC GTAGATACGG TGACAGAAAAG	7020
CAAGATTCAG CATGCCATGG AGGTGGTTGT AGCAGGTAGA ACTAGTTTCG TCATTGCCCA	7080
CCGCTTGAAA ACCATTCTCA ATGCAGATCA GATTATTGTC CTTAAAGATG GAGAAATCAT	7140
TGAACGTGGT AACCACCATG AACTTTTGAA GCTAGGTGGC TTTTATTTCAG AACTCTATCA	7200
CAATCAATTT GTTTTCGAAT AAGAAAGAAG TTGTCCTATG TGGGCAGCTT TTTCTTGTCC	7260
ATAAAAAATG TTTATCACAG CCTTAAAAAA AACATATTAG ACGAAAGTCA TTTTGAGTGA	7320
TATGATAGGA CTATCGTTAG CATTGGAAG GAGAGGCATC ATGGCTAGAA CGGTGTAGG	7380
AGTTGCTGCA AATCTATGTC CCGTAGACGC AGAAGGCCAA ATCATTTCATT CATCTGTATC	7440

537

TTGTAGATTC GCAGAGATCA TTCGTCAAGT CGGTGGTCTC CCTTTAGTCA TTCCTGTTGG	7500
TGATGAGTCA GTTGTACGTG ATTATGTGGA AATGATTGAC AAACTCATTT TGACAGGAGG	7560
CCAAAATGTT CATCCTCAGT TTTATGGAGA GAAAAAGACC GTCGAGAGCG ATGATTACAA	7620
TCTGGTCCGT GACGAATTTG AATTGGCACT CTTGAAGGAA GCGCTTCGTC AGAATAAACC	7680
AATTATGGCA ATCTGTCGCG GTGTCCAAC TGTCAATGTT GCCTTTGGTG GAACCCCTCAA	7740
TCAAGAAATC GAAGGTCAGG	7760

(2) INFORMATION FOR SEQ ID NO: 64:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2723 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GAGGTTTTAA TTCACTTACC TCTSCCGTAT CTTTATTTAA AATGAATTCT TTTACGGTTG	60
TATTTCTTGC AAAATCTTTT ACAACAATCT TAATGTTTAG TGTCTTGTCT ATTATTGTTT	120
TAATATCATT AAATGATGTA TATCTTTTC CATTATATA AATATGTTGT TCTTGAATCT	180
CACCATCGAA TCCATTATTT CTTTATCAT TGATGTTAAA GACTACAGAT TTTCCATCAG	240
CATATTCGAT ACTAGTATTT CCCTTAGGAT CAATGTTTAC TTCGGGTTTA ACATTATCAT	300
ATAAAAACTG ATAGTGGACT CCAACTGCTT TAGCATTCAA ATCGCTATAG CCAGTTTGAA	360
GATAAACATT TCCATCCATA TCTGTTACCT TATCTGGAAA TCCGTTTGCT TTATAGTCTT	420
TCATTCCCCA GTCCATGATG TCACCGTCTT TAACATTCAG CTTAATATTA AAATCTCTAG	480
TGTTATCAAT GTGTAAATCT CCGTAGATTA AATAATTATC TACAACCGAT TCATTAACTC	540
TCAATTCCCC GTTAAAACCA CCCTTATCAG AAATCTTACC TETTAAATAA AATTCTGGAT	600
TTCGTACATA AATTTTATTA GATTTAGATG GATTAAAGTA GTTCTTATCC ATTGAAAGGT	660
TTACTGGTTT GGTATCAATA AATAACATGG AGCCATCTTC TTTTATAGCT TCTACATTGA	720
ACTTATCCTC TCCAGTGTAT TCTTTATCAT CTTACCAAAA TAATACAAGT TTAGAAGAAT	780
CTGTCACAAG ATTTCCGTCT TTATCGATAG CTTCCCTTTT ATCGTTCATT TTAAATGTAA	840
ACACTTGATA CCTTATAATG TTAAAGCCGT CCAAAGCCGA CATTAAATACA GATTGGGTAC	900
TTCTTCCATC TTCAACATTT CTAATATCAG CATAAATTGT TGTTTCTGAA AGGGCTCTTA	960
GATTAGGATT GGCCTTTGT ATTTTGTGTA TATCTTCCTT GCTATAGACT CCATTTCTT	1020

538

CTAACATATC CGTTTTCCTCA GGATTATAGG TAGTCACCTT TAGTGCATAG CCTTTCTTA	1080
GAATGATATT ATCCTTTAAC AGATATTGTT GTTTTCTGA ATCAGAATAG ATTTTACCAG	1140
ATTCATTTT AGTTAAATTG TCTGGTTTGT TTTTGAAAG ATCTCCTCC CCTAATTCTA	1200
TGACATTCCC ATAACCTGAT ACATAGGGAT ATTCTGATT AGTTTCCTTA ATTTTTCAG	1260
GCATTCTAAT TTTAATTTCA GCTTTTCT GATCATTATC TTTAACAAAT AATCTCATAT	1320
CTCCTGCAAA AGCTAATCCA TCCACAATAT CATTAATATT AGCGTATAGA TCAAATGTCA	1380
TCGTTTGTGA GTGGAAATCA TACTGGTCG CTTTGATTTC TATAGATTTA TAGTTATTCC	1440
CATAATATAC CTTGGCATTT TTAGAAACAT TACTTATCTT TCCAAGAATT TCAAAGTGTC	1500
CATCTTTAGA CGGACTTAGA ACACCATAAA TTTTGATTT GATTTCGTCA AGTTTCTCAG	1560
TTTCATATTC TAGATCAGTC CCATCATCGT AGGCTATTAT ATTTCTTTA TCATCGTATT	1620
TATAATCGTA TTCCTCCATT CTCTTACCAG TTTCACTTGT AAAATCATCA ACTTCTCTAA	1680
ATTTCTTTT AATGAGTTTC TTTAAGTCTT TATTTTCAA GTCTCTAATT GTTGAAATAT	1740
TTCTATCAAT AGTAAACTA GATTTTCTT TAATAGACTC TTCATTTCT TGATGATGAT	1800
GTCTACCCC AGTTGTATCT TTTTGTAGAC TACCCTCTT TCCATTTCT AAATTTTAA	1860
ATTTAGATTC TGCAATCTCG CCAAGCTTT GATATTTAGA TGAATCTGA TCAGGATCTA	1920
CTAGATAATA GGAAATCAT CCCTTTTCAT CAGCCTGATT AGCAAATTTA ATTCTATGAA	1980
TCTTTGTGAA ATTGCTAGAA CCATCTAATG CAATGACTTC AATGATTTT CCCCTTAAAT	2040
CTCCCGCACC TTTAATTTCA TAAATGGTAT TTCCGTCTT ATCAAGTTT CTATTCTTC	2100
CTTGACCCTC ACCTGCGTAA GTTACTTCAA GATTTTTC AACCTCTCCA TCTTCATTAA	2160
CAAGAGCGGC GCCAGCATAC CAAACTTCGT TCGCAATCTC GTCAAATTT TCAGGATGTT	2220
CTTTTGTATC TCTCGCAAAT AGCGTTTCAT TCTTATACTG ATCTTTTACC TTATGATAAG	2280
TATCCTTTGT AATCAACTTA ATTTTTCAG GATTTGAAAA ATCAACCGAA ACAATCTTAG	2340
GGGCGGTGTT ATCAATTTT ACAGGAATAT AGGAAACCTG CCATGGGTAA TCTTTAGTTA	2400
ATCTATATTT AAATTTATAG AAATATTGAC CTTCGCAAT CGGTTCAAAT TGACCTCTTA	2460
TCTTAGTAGC AGGATCTTGA TTATCCTTAC TTTCTGGTGC ATTTCTTCT CTACCTCTAG	2520
GATTATAGAT GAGTCCATCC CACTTCAAGT CACCCCAAAC TTTAGTTTA GATGATTGA	2580
TTCCCTTTC ATCATTGCTT TTAGAAATTA AAATTCCTCT AATAAAGTGT TCTCTGAAA	2640
TGACTTTTAA GTCTCTTTGA TTTCTCCCT CTTATTGTT ATTTACTATT GAAATCAATC	2700
CTTCTTCTGC ACTTCTTAAT ACA	2723

(2) INFORMATION FOR SEQ ID NO: 55:



539

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11811 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

AAAAAAGTGG GAATGACTCA AATCTTCACT GAAGCTGGCG AATTGATCCC TGTAACAGTT	60
ATTGAAGCAA CTCCAAACGT TGTCTTCAA GTTAAACTG TTGAAACAGA CGGATACAAC	120
GCTATCCAAG TTGGTTTCGA TGACAAACGC GAAGTATTGA GCAACAAACC TGCTAAAGGA	180
CATGTAGCGA AAGCTAACAC GGCTCCTAAG CGCTTCATTC GTGAATTCAA AAACGTTGAA	240
GGCTTGGAAG TTGGTGCTGA AATTACAGTT GAAACATTCG CAGCTGGAGA CGTTGTTGAC	300
GTAACGGGTA CTTCTAAAGG TAAAGGTTTC CAAGGTGTTA TCAAACGCCA CGGACAATCA	360
CGTGGACCAA TGGCTCACGG TTCTCGTTAC CACCGTCGTC CAGGTTCTAT GGGGCCTGTT	420
GCACCTAACC GCGTATTCAA AGGTAAAAAC CTTGCAGGAC GTATGGGTGG CGACCGCGTA	480
ACAATTCAAA ACCTTGAAGT TGTACAAGTT GTTCCAGAAA AGAACGTTAT CCTTATCAAA	540
GGTAACGTAC CAGGTGCTAA GAAATCTCTT ATCACTATCA AATCAGCAGT TAAAGCTGGT	600
AAATAATAAA GAAAGGGGAA ATCAGTCACA ATGGCAAACG TAACATTATT TGACCAAACCT	660
GGTAAAGAAG CTGGCCAAGT TGTCTTAGC GATGCAGTAT TTGGTATCGA ACCAAATGAA	720
TCAGTTGTGT TTGATGTAAT CATCAGCCAA CGCGCAAGCC TTCGTCAAGG AACACACGCT	780
GTAAAAAACC GCTCTGCAGT ATCAGGTGGT GGACGCCAAC CATGGCGTCA AAAAGGAACT	840
GGACGTGCTC GTCAAGGTTG TATCCGCTCA CCACAATGGC GTGGTGCTGG TGTGTCTTTC	900
GGACCAACTC CACGTTTATA CGGCTACAAA CTTCCACAAA AAGTTCGTGG CTTAGCTCTT	960
AAATCAGTTT ACTCTGAAAA AGTTGCTGAA AACAAATTCG TAGCTGTAGA CGCTCTTTCA	1020
TTTACAGCTC CAAAACTGC TGAATTTGCA AAAGTTCTTG CAGCATTGAG CATCGATTCT	1080
AAAGTTCTTG TTATCCTTGA AGAAGGAAAT GAATTCGCAG CTCTTTCAGC TCGTAACCTT	1140
CCAAACGTGA AAGTTGCAAC TGCTACAACCT GCAAGTGTTT TTGACATCGC AAATAGCCAC	1200
AAACTTCTTG TCACACAAGC AGCTATCTCT AAAATCGAGG AGGTTCTTGC ATAATGAATT	1260
TGTATGATGT TATCAAAAAA CCTGTCATCA CTGAAAGCTC AATGGCTCAA CTTGAAGCAG	1320
GAAAAATATGT ATTTGAAGTT GACACTCGTG CACACAAACT TTTGATCAAG CAAGCTGTTG	1380
AAGCTGCTTT CGAAGGTCTT AAAGTTGCCA ATGTTAACAC AATCAACGTA AAACCAAAAG	1440

540

CTAAACGTGT TGGACGTTAC ACTGGTTTTA CTAACAAAAC TAAAAAAGCT ATCATCACAC	1500
TTACAGCTGA TTCTAAAGCA ATCGAGTTGT TTGCTGCTGA AGCTGAATAA TCTAAGGAGG	1560
AAATATCGTG GGAATTCGTG TTTATAAACC AACAACAAAC GGTGCGCGTA ATATGACTTC	1620
TTTGGATTTC GCTGAAATCA CAACAAGCAC TCCTGAAAAA TCATTGCTTG TTGCATTGAA	1680
GAGCAAGGCT GGTCGTAACA ACAACGGTCG TATCACAGTT CGTCACCAAG GTGGTGGACA	1740
CAAACGTTTC TACCGTTTGG TTGACTTCAA ACGTAATAAA GACAACGTTG AAGCAGTTGT	1800
TAAAACAATC GAGTACGATC CAAACCGTTC TGCAACATC GCTCTTGATC ACTACACTGA	1860
CGGTGTGAAA GCATACATCA TCGCTCCAAA AGGTCTTGAA GTAGGTCAAC GTATCGTTTC	1920
AGGTCCAGAA GCAGATATCA AAGTCGAAA CGCTCTTCCA CTTGCTAACA TCCCAGTTGG	1980
TACTTTGATT CACAACATCG AGTTGAAACC AGGTGCTGGT GGTGAATTGG TACGTGCTGC	2040
TGGTGCACTC GCTCAAGTAT TGGGTTCTGA AGGTAAATAT GTTCTTGTTT GTCTTCAATC	2100
AGGTGAAGTT CGTATGATTC TTGGAACCTG CCGTGCTACA GTTGGTGTG TCGGAAACGA	2160
ACAACATGGA CTTGTAAACC TTGGTAAAGC AGGACGTAGC CGTTGGAAAG GTATCCGCCC	2220
AACAGTTCGT GGTTCGTGTA TGAACCTTAA CGATCACCCA CACGGTGGTG GTGAAGGTAA	2280
AGCACCAGTT GGTCGTAAAG CACCATCTAC TCCATGGGGC AAACCTGCTC TTGGTCTTAA	2340
AACTCGTAAC AAGAAAGCGA AATCTGACAA ACTTATCGTT CGTCGTCGCA ACGAGAAATA	2400
ATATTAAACT AGTCGCTTAA GCAACTAGTA AATCCGCCAG CTCGGTAGCG CTCCATAGGA	2460
GTGCAAGCCG CTGTGGTACA ACATTTAAAG GAGAAAATAT AAAAATGGGA CGCAGTCTTA	2520
AAAAAGGACC TTTCGTCGAT GAGCATTTGA TGAAAAAGT TGAAGCTCAA GCTAACGACG	2580
AAAAGAAAA AGTTATTAAA ACTTGGTCAC GTCGTTCAAC GATCTTCCCA AGTTTCATTG	2640
GTTACACTAT TGCAGTTTAT GACGGACGTA AACACGTACC TGTTTACATC CAAGAAGACA	2700
TGGTAGGCCA CAAACTTGGT GAATTTGCAC CAACTCGTAC TTACAAAGGT CACGCTGCAG	2760
ACGACAAGAA AACACGTAGA AAATAAGGAG AACATAAATG GCAGAAATTA CTTCAGCTAA	2820
AGCAATGGCT CGTACAGTAC GTGTTTCACC TCGTAAATCA CGTCTTGTTT TTGATAACAT	2880
CCGTGGTAAA AGCGTAGCCG ATGCAATCGC AATCTTGACA TTCACTCCAA ACAAAGCTGC	2940
TGAAATCATC TTGAAAGTTT TGAATCAGC TGTAGCTAAC GCTGAAAACA ACTTTGGTTT	3000
GGATAAAGCT AACTTGGTAG TATCTGAAGC ATTGCGAAAC GAAGGACCAA CTATGAAACG	3060
TTTCCGTCCA CGTGCAGAAAG GTTCAGCTTC ACCAATCAAC AAACGTACAG CTCACATCAC	3120
TGTAGCTGTT GCAGAAAAAT AAGGAGGTAA AATCGTGGGT CAAAAAGTAC ATCCAATTGG	3180
TATGCGTGTC GGCATCATCC GTGATTGGGA TGCCAAATGG TATGCTGAAA AAGAATACGC	3240

541

GGATTACCTT CATGAAGATC TTGCAATCCG TAAATTCGTT CAAAAAGAAC TTGCTGACGC	3300
AGCAGTTTCA ACTATTGAAA TCGAACGCGC AGTAAACAAA GTTAACGTTT CACTTCACAC	3360
TGCTAAACCA GGTATGGTTA TCGGTAAAGG TGGTGCTAAC GTTGATGCAC TCCGTGCAAA	3420
ACTTAACAAA TTGACTGGAA AACAAGTACA CATCAACATC ATCGAAATCA AACAACTGA	3480
TTTGGATGCT CACCTTGTAG GTGAAGGAAT TGCTCGTCAA TTGGAGCAAC GTGTTGCTTT	3540
CCGTCGTGCA CAAAAACAAG CAATCCAACG TGCAATGCGT GCTGGAGCTA AAGGAATCAA	3600
AACTCAAGTA TCAGGTCGTT TGAACGGTGC AGATATCGCC CGTGCTGAAG GATACTCTGA	3660
AGGAACTGTT CCGCTTCACA CACTTCGTGC AGATATCGAT TACGCTTGGG AAGAAGCAGA	3720
TACTACATAC GGTAACCTTG GTGTTAAAGT ATGGATCTAC CGTGGTGAAG TTCTTCCAGC	3780
TCGTAAAAAC ACTAAAGGAG GTAAATAACC AATGTTAGTA CCTAACGTG TTAACACCG	3840
TCGTGAGTTC CGTGAAAAA TGC CGGTGA AGCAAAAGGT GGAAAAGAAG TAGCATTCCG	3900
TGAATACGGT CTTCAAGCTA CAACTAGCCA CTGGATCACT AACCGCCAAA TCGAAGCTGC	3960
TCGTATCGCC ATGACTCGTT ACATGAAACG TGGTGGTAAA GTTTGGATTA AAATCTTCCC	4020
ACACAAATCA TACTACTGTA AAGCTATCGG TGTGCGTATG GGATCTGGTA AAGGGGCACC	4080
TGAAGGTTGG GTAGCACCAG TTAACGTGG TAAAGTGATG TTCGAAATCG CTGGTGTATC	4140
TGAAGAGATT GCACGTGAAG CGCTTCGACT TGCTAGCCAC AAATTGCCAG TTAATGTAA	4200
ATTCGTAAAA CGTGAAGCAG AATAAGGAGA AGGCATGAAA CTTAATGAAG TAAAAGAATT	4260
TGTTAAAGAA CTTCTGGTTC TTTCTCAAGA AGAACTCGCG AAGCGCGAAA ACGAATTGAA	4320
AAAAGAATTG TTTGAACCTC GTTTCCAAGC TGCTACTGGT CAATTGGAAC AAACAGCTCG	4380
CTTGAAAGAA GTTAAAAAAC AAATCGCTCG CATCAAAACA GTTCAATCTG AAGCGAAATA	4440
ATAGACTAGG GAAGGAGAAA TTTCAATGGA ACGCAATAAT CGTAAAGTTC TTGTTGGACG	4500
TGTTGTATCT GACAAAATGG ACAAGACAA CACAGTTGTA GTTGAAACAA AACGTAACCA	4560
CCCAGTCTAT GGTAACGTA TTAAGTCTC TAAAAAATAC AAAGCTCATG ATGAAAACAA	4620
TGTTGCCAAA GAAGCGGATA TCGTACGTAT CATGGAAACT CGCCCGCTTT CAGCTACAAA	4680
ACGTTTCCGT CTTGTAGAAG TTGTTGAAGA AGCGGTCATC ATCTAATCAA ACCTGAAAGG	4740
AGAAAACTGA AATGATTCAA ACAGAACTC GTTTGAAAGT CGCAGACAAC AGCGGTGCTC	4800
GCGAAATCTT GACTATCAAA GTTCTTGGTG GTTCAGGACG TAAATTGCA AACATCGGTG	4860
ATGTTATCGT GGCATCTGTA AAACAAGCTA CTCCTGGTGG TCGCGTTAAA AAAGGTGACG	4920
TTGTTAAAGC AGTTATCGTT CGTACTAAAT CAGGTGCTCG TCGTCTGAT GGTTCATACA	4980

542

TCAAATTTGA CGAAAACGCA GCAGTTATCA TCCGTGAAGA CAAAACTCCT CGCGGAACAC	5040
GTATCTTTGG CCCAGTTGCA CGTGAATTGC GTGAAGGTGG CTTATGAAG ATCGTGTAC	5100
TTGCTCCAGA AGTACTTTAA TTTTATAGGAA CAACTAGTC CCCTAGCTTC AAGCTAGGGT	5160
CCCCTTATGG GCGTAAGAAA AATCAAGGAG AAACCTAATG TTTGTAAAAA AAGGCGACAA	5220
AGTTCGCGTA ATCGCTGGTA AAGATAAGGG AACAGAAGCT GTTGTCTTA CTGCCCTTCC	5280
AAAAGTAAAC AAAGTTATCG TTGAAGGTGT TAACATTGTT AAGAAACACC AACGTCCAAC	5340
TAACGAGCTT CCTCAAGGTG GTATCATCGA GAAAGAAGCA GCTATCCACG TATCAAACGT	5400
TCAAGTTTTG GACAAAAATG GTGTAGCTGG TCGTGTGGA TACAAATTTG TAGACGTTAA	5460
AAAAGTTCGC TACAACAAAA AATCAGGCGA AGTGCTTGAT TAATCAGGAA GGAAAGGAGA	5520
AGTATAATGG CAAATCGTTT AAAAGAAAAA TATCTTAATG AAGTAGTTCC TGCTTTGACA	5580
GAACAATTTCA ACTACTCATC AGTGATGGCT GTGCCTAAAG TAGATAAGAT TGTTTTGAAC	5640
ATGGGTGTTG GTGAAGCTGT ATCAAACGCT AAAAGCCTTG AAAAAGCTGC TGAAGAATTG	5700
GCACCTATCT CAGGTCAAAA ACCACTTATC ACTAAAGCTA AAAATCAAT CGCCGGCTTC	5760
CGTCTTCGTG AAGGTGTTGC GATCGGTGCA AAAGTTACCC TTCGTGGTGA ACGTATGTAC	5820
GAATTCCTGG ATAAATTGGT ATCAGTTTCA CTTCCACGTG TACGTGACTT CCACGGTGTC	5880
CCAACAAAAT CATTTGATGG ACGCGGGAAC TACACACTTG GTGTGAAAGA ACAATTATC	5940
TTCCAGAAA TCAACTTCGA TGACGTTGAC AAAACTCGTG GTCTTGACAT CGTTATCGTA	6000
ACAACTGCTA AACTGACGA AGAGTCACGT GCATTGCTTA CAGGCCTTGG AATGCCTTTT	6060
GCAAAATAAT ATAGGAGGTA AATCTAATGG CTAAAAATC AATGGTACCT AGAGAGGCTA	6120
AACGCCAAAA AATTGTTGAC CGTTATGCTG AAAACCTGC TCCATTAAAG GCGGCAGGGG	6180
ACTACGAAGG TTTATCTAAA TTACCTCGCA ACGCCTCACC GACTCGTTTA CATAATCGTT	6240
GTAGGGTTAC GGGGCGCCCA CATTCAAGTT ACCGCAAAT TGGTCTGACT CGTATCGCTT	6300
TTCCGGAAT TGCGCATAAA GGTCAAATTC CTGGTGTAAC AAAAGCATCT TGCTAATTTA	6360
AGATATCAAG AGCGTCAAAA CTCCAAGTAA AAATAGGAAA CTTGACGAAG AACTAAAGT	6420
TTCTAGGAAA GTTTATCTTT TTCACACAGA GTTTAGCCCG GGTCAATTG GGCTTGCCAA	6480
TTTGAACACG AGCTACAGCT TTGGCAAAAA AGACCAATTT GCTTTGGAGC ATTGCTTCTG	6540
CATTAAATTG TCTATTTTTC CTCGTGCTGT TACGCTCTT GTATCATGTA TTAAGTAGCA	6600
AGTGCAACTT GCAAACTACT AGTAAGAGGA GAAAAACAAA ATCGTTATGA CTGACCCAAT	6660
CGCAGACTTC CTAAGTCGTA TTCGTAATGC TAACCAAGCT AAACACGAAG TACTTGAAGT	6720
ACCTGCATCA AACATCAAAA AAGGGATTGC TGAATCCTT AAACCGGAAG GTTTTGTAAG	6780

543

AAACGTTGAA ATCATTGAAG ATGACAAACA AGGCGTCATC CGTGTATTTT TAAATACGG	6840
ACCAAAATGGT GAGAAAGTTA TCACTAACTT GAAACGTGTT TCTAAACCAG GACTTCGTGT	6900
CTACAAAAAA CGTGAAGACC TTCCAAAAGT TCTTAACGGA CTTGGAATTG CCATCCTTTC	6960
AACCTTCTGAA GTTTTGCTTA CTGATAAAGA AGCACGCCAA AAGAATGTTG GTGGTGAGGT	7020
TATCGCTTAC GTTTGGTAAA ATCAAGATAC AAAGCTCGTA AAGAACAAAG CAAAATTAGG	7080
AAGTTGGAGA AGTTTGTTTA CAAACAAGCC AACTTATCTA TTTTGCACAG TTCTTAGAGC	7140
GTGTTCAAGT CAGCTCTTGA ACTAAATAAG TATCTGAACC CCGTAAAAAC TGGCCGTCT	7200
GGCCTGACAA TTTAACAGGA GAAAATAAAC ATGTCACGTA TCGGTAATAA AGTTATCGTG	7260
TTGCCTGCTG GTGTTGAAGT CGCTAACAAT GACAACGTTG TAACTGTAAA AGGATCTAAA	7320
GGAGAACTTA CTCGTGAGTT CTCAAAAGAT ATTGAAATCC GTGTGGAAGG TACTGAAATA	7380
ACTCTTCACC GTCCAAACGA TTCAAAGAA ATGAAAACTA TCCACGGAAC TACTCGTGCC	7440
CTTTTGAACA ACATGGTTGT TGGTGTATCA GAAGGATTCA AGAAAGAACT TGAATGCGT	7500
GGGGTTGGTT ACCGTGCACA GCTTCAAGGA TCTAAACTTG TTTTGGCTGT TGGTAAATCT	7560
CATCCAGACG AAGTTGAAGC TCCAGAAGGA ATTACTTTTG AACTTCCAAA CCCAACAACA	7620
ATCGTTGTTA GCGGAATTTT AAAAGAAGTA GTTGGTCAAA CAGCTGCTTA CGTACGTAGC	7680
CTTCGTTTAC CAGAACCATA TAAAGGTAAA GGTATCCGTT ACGTTGGTGA ATTCGTTCCG	7740
CGTAAAGAAG GTAAAACAGG TAAATAATGT TGAGTGGTTG ATCATCAACC ACCAACCAT	7800
TTTCCAACCT TGTGCATAGC ACACGATTTA AACTAAAGA GGTGAAAACT GTGATTTCAA	7860
AACCAGATAA AAACAAACTC CGCCAAAAAC GCCACCGTCG CGTTCGCGGA AACTCTCTG	7920
GAACCTGCTG TCGCCACGTT TGAACGTAT TCCGTTCTAA TACAGGCATC TACGCTCAAG	7980
TGATTGATGA CGTAGCGGGT GTAACGCTCG CAAGTGCTTC AACTCTTGAT AAAGAAGTTT	8040
CAAAAGGAAC TAAACTGAA CAAGCCGTTG CTGTCGGTAA ACTCGTTGCA GAACGTGCPA	8100
ACGCTAAAGG TATTTTCAAG GTGGTGTTTG ACCGCGGTGG ATATCTATAT CACGGACGTG	8160
TGAAAGCTTT GGCTGATGCA GCTCGTGAAA ACGGATTGAA ATTCTAATAG GAGGACACTA	8220
GAAAATGGCA TTTAAAGACA ATGCAGTTGA ATTAGAAGAA CGCGTAGTTG CTGTCAACCG	8280
TGTTACAAAA GTTGTTAAAG GTGGACGTCG TCTTCGTTTC GCAGCTCTTG TTGTTGTTGG	8340
TGACCACAAT GGTGCGGTAG GATTGCTTAC TGGTAAAGCT CAAGAAGTTC CAGAAGCAAT	8400
CCGTAAAGCA GTAGATGATG CTAAGAAAAA CTTGATCGAA GTTCCTATGG TTGGAACAAC	8460
AATCCACAC GAAGTTCTTT CAGAATTCGG TGGAGCTAAA GTATTGTTGA AACCTGCTGT	8520

544

AGAAGGTTCT GGAGTTGCCG CTGGTGGTGC AGTTCGTCCC GTTGTGGAAT TGGCAGGTGT	8580
GGCAGATATT ACATCTAAAT CACTTGGTTC TAACACTCCA ATCAACATTG TTCGTGCAAC	8640
TGTTGAAGGT TTGAAACAAT TGAAACGCGC TGAAGAAATT GCTGCCCTTC GTGGTATTTC	8700
AGTTTCTGAT TTGGCATAAG AAAGGGGATA AAATGGCTCA AATTAAAATT ACTTTGACTA	8760
AGTCTCCAAT CGGACGCATT CCATCACAAC GTAAACTGT TGTAGCACTT GGACTTGGCA	8820
AATTGAACAG CTCTGTTATT AAAGAAGATA ACGCTGCTAT CCGTGGTATG ATCACAGCAG	8880
TATCTCACTT AGTAACAGTT GAAGAAGTAA ACTAATGAAG TTTTAGGGGA TGTGCACTGT	8940
ACCATCCCTT AAAACTAGAT ATAGTCATCT ATGATGACAT CGTATAGGCG AGTTGATGGG	9000
GGAGACAACC TTTTCTCCCT TATCGGCCCT AGCATTTTAC AAAAGAGGAG AAAATAAAAA	9060
TGAAACTTCA TGAATTGAAA CCTGCAGAAG GTTCTCGTAA AGTACGTAAC CGCGTTGGTC	9120
GTGGTACTTC ATCAGGTAAC GGTA AACAT CTGGTCGTGG TCAAAAAGGT CAAAAGCTC	9180
GTAGCGGTGG CGGAGTTGCG CTGGTTTTG AAGGTGGACA AACTCCATTG TTCGTCGTC	9240
TTCCAAAACG TGGATTCACT AACATCAACG CTAAAGAATA CGCAATTGTG AACCTTGACC	9300
AATTGAACCT CTTTGAAGAT GGTGCTGAAG TAACTCCACT TGTCTTATC GAAGCAGGAA	9360
TTGTTAAAGC TGAAAAGTCA GGTATTAAAA TTCTTGGTAA CGGTGAGTTG ACTAAGAAAT	9420
TGACTGTGAA AGCAGCTAAA TTCTCTAAAT CAGCTGAAGA AGCTATCACT GCTAAAGGTG	9480
GTTCAGTAGA AGTCATCTAA GAGAGGTGAC CTATGTTTTT TAAATTATTA AGAGAAGCTC	9540
TTAAAGTCAA GCAGGTTCCA TCAAAAATTT TATTTACAA TTTTATCGTT TTGGTCTTTC	9600
GTATCGGAAC TAGCATTACA GTTCTGGTG TGAATCCCA TAGCTTGAAT GCTTTAAGTC	9660
GATTATCCTT CTTAAACATG TTGAGCTTGG TGTCGGGGAA TGCCCTAAAA AACTTTTCCA	9720
TTTTTGCCCT AGGAGTTAGT CCTATATCA CCGCTTCTAT TGTGTCCAA CTCTTGCAAA	9780
TGGATATTTT ACCCAAGTTT GTAGAGTGGG GTAAACAAGG GGAAGTAGGT CGAAGAAAAT	9840
TGAATCAAGC TACTCGTTAT ATTGCTCTAG TTCTCGCTTT TGTGCAATCT ATCGGGATTA	9900
CAGCTGGTTT TAATACCTTG GCTGGAGCTC AATTGATTAA AACTGCTTTA ACTCCACAAG	9960
TTTTTCTGAC GATTGGTATC ATCTTAACAG CTGGTAGTAT GATTGTCACT TGGTTGGGTG	10020
AGCAAATTAC AGATAAGGGA TACGGAAACG GTGTTTCCAT GATTATCTTT GCCGGGATTG	10080
TTTCCTCAAT TCCAGAGATG ATTCAGGGCA TCTATGTGGA CTACTTTGTG AACGTCCCAA	10140
GTAGCCGTAT CACTTCATCT ATCATTTTCG TAATCATTTT GATTATTACT GTATTGTTGA	10200
TTATTACTT TACAACCTAT GTTCAACAAG CAGAATACAA AATTCCAATC CAATATACTA	10260
AGGTTGCACA AGGTGCTCCA TCTAGCTCTT ACCTTCGTT AAAAGTAAAC CCTGCTGGAG	10320

545

```

TTATCCCTGT TATCTTTGCC AGTTCGATTA CTGCAGCGTG CGGCTATTCT TCAGTTTTTG 10380
AGTGCCACAG GTCATGATTG GGCTTGGGTA AGGGTAGCAC AAGACATGTT GGCAACTACT 10440
TCTCCAAC TGATTGCCAT GTATGCTTTG TTGATTATTC TCTTTACATT CTCTATACG 10500
TTTGTACAGA TTAATCCTGA AAAAGCAGCA GAGACCTAC AAAAGAGTGG TGCCTATATC 10560
CATGGAGTTC GTCCTGGTAA AGGTACAGAA GAATATATGT CTAAACTTCT TCGTCGTCTT 10620
GCAACTGTTG GTTCCCTCTT CCTTGGTGTG ATTTCCATTT TACCGATTGC AGCTAAAGAT 10680
GTATTTGGTC TTTCTGATGT TGTTCGCTTT GGTGGAACAA GTCTCTTGAT CATTATCTCT 10740
ACAGGTATCG AAGGAATCAA GCAATTGGAA GGTACCTAT TGAAACGTAA GTATGTTGGT 10800
TTCATGGACA GAACAGAATA AAAGTATTTA CTGAATCACT AAATACTGAG GGAGTGGAGG 10860
TTTAAACTCT GACATTTGTA AGAGTTGGAT CTCCCTCTT CTATTTTGT TTTAAATCGG 10920
GGTGAAAAGA CTTTTTGCTT CTATTTAAAA ATAAATAAG GAGATCAAAT CATGAATCTT 10980
TTGATTATGG GCTTACCTGG TGCAGGTAAG GGAACCAAG CAGCAAAAAT CGTAGAACAA 11040
TTCCATGTTG CACATATCTC AACAGGTGAT ATGTTCCGCG CTGCAATGGC AAATCAAAC 11100
GAAATGGGTG TTCTTGCTAA GTCATATATT GACAAGGCTG AATTGGTTCC TGACGAAGTT 11160
ACAAATGGAA TCGTAAAGA ACGCCTTTCA CAAGATGATA TTAAAGAAAC AGGATTCTTA 11220
TTGGATGGTT ACCCACGTAC AATTGAACAA GTCATGCCT TGGACAAAAC ATTGGCTGAA 11280
CTTGGCATTG AACTAGAAGG TGTTATCAAT ATTGAAGTA ACCCTGACAG CCTTTTGAA 11340
CGTTTGAGTG GCGTATCAT CCACCGCGTA ACTGGAGAAA CTTTCCACAA GGTCTTTAAC 11400
CCACCAGTTG ACTATAAGA AGAAGATTAC TACCAACGTG AAGATGATAA GCCTGAGACA 11460
GTAAAACGTC GTTTGGATGT TAATATTGCT CAAGGAGAAC CAATCATTGC TCACTACCCT 11520
GCCAAAGGTT TGGTTCATGA CATCGAAGGT AATCAAGATA TCAATGATGT CTCTCAGAT 11580
ATTGAAAAAG TATTGACAAA TTTGAAATAA AGCGTTTTTC ACACTTGCAA AAATCCGCTA 11640
CAAATGTTAT ACTGAGATAG TCTGACTTAT AATTGTTGTC TCTGTCTTA GAGGCATCGA 11700
ATCGAAATTT ATGGAGGTGC TTTTGGTGG CAAAAGACGA TGTGATTGAA GTTGAAGGCA 11760
AAGTAGTTGA TACAATGCCG AATGCAATGT TTACGGTTGA ACTTGAAAAT GGACATCAGA 11820
TTTTAGCAGG G 11831

```

(2) INFORMATION FOR SEQ ID NO: 66:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10726 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

546

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

CCCCGCATTT GAAAGCTATT CGTGAAGGAT TTATGATGGC AATGCCTTTG ATTTTAGTCG	60
GCTCTTTATT TCTTATTCTA ATCAGTTGGC CTCAAGAGGC TTTTACAAAT TGGCTGAATA	120
GTGTTGGATT GCTAAGTATC TTGACAACTA TGAATCAGTC AACAGTAGCG ATTATCTCCT	180
TGGTCGCTTG TTTCGGTATT GCCTACAGGT TGTCGGAAGG ATATGGTACA GATGGTCCGT	240
CGGCAGGGAT CATAGCCTTA TCCAGTTTIG TATTGATGGC ACCTCGTTTT TCGAGTATGG	300
TTTATGATAA AAATGGGGAG CAGGTCAAGC AGTTATTTGG CGGCGCAATA CCATTTTCTA	360
GCCTGAATGC ATCTTCTTTG TTTATGGCGA TTAATTTGG ATTGGTTACA GCAGAGATTT	420
ATCGTATGTT TATCCAGCGC GGAATTACGA TAAAAATGCC AAGTGGTGTC CCAGATGTAG	480
TAAGTAAATC ATTTTCAGCT CTTTTATCTG GTTTTACTAC TTTTGTTTTG TGGGCTTTGG	540
TCTTAAAAGG TCTTGAAGCG GCAGGAGTTG CAGGAGGTCT CAACGGAATC CTAGGTGCAA	600
TTGTTGGAAC ACCGCTTAAG TTAATTGCAG GAACGCTTCC AGGTATGATT CTATGTGTTA	660
TTGTAAATC ATTCTTTTGG TTCTGTGGAG TTAATGGGGG ACAAGTTTTA AATGCTTTTG	720
TAGACCCAGT TTGGTTACAA TTTACTACAG AAAACCAAGA AGCTGTGGCT GCAGGACAAA	780
CACTCCAACA CATTATTACA TTACCGTTTA AACATTTATT TGTATTTATT GGTGGCGGTG	840
GAGCGACTAT TGGTCTTGGC ATTTGTCTCT TCCTATTTAG TAAGAGTCGT GCGAATAAAA	900
CATTAGGTAA GCTAGCTATT ATACCGTCTA TTTTAAATAT CAATACAGCT ATTCTATTTA	960
CGTTTCCAAC AGTTTTAAAT CCGATTATGC TGATTCCGTT TATTGCTACT CCTACAATCA	1020
ATGCCCTGAT TACCTATGTA TCAATGGCTG TAGGATTAGT ACCCTATACA ACAGGTGTAA	1080
TCCTTCCGTG GACAATGCCA CCGATTATAG GAGGCTTCCT TGCAACAGGG GCTAGTTGGC	1140
GAGGAGCTCT ATTACAAGTT GTTTTGATTT TGGTTTCTGT AGCAATTTAT TATCCATTCT	1200
TCAAAATTGC AGATAAACGC AATCTTGAAA AAGAAAAAGC TACTGTTGGA GGGAAATAAG	1260
ATGGTTATCA GAGTATTTGA TCAACAGAAA AATACTTATT CTAGCTTTGC CTTAGAGGAA	1320
TTAAGTTACT ATATGAATCG GGTCTTTAAG ACTAACATAG AGCTTGTCGA GGAGAAGGAA	1380
GCGGATATTT TTGTAGGATT AGTCAATAAA GAGGACAGAA AAGACCATGT TCTTATCTCA	1440
TTAGACAAGG GTAAGGGGAG AATTGAGTCT AATACAATTG TAGGTTTACT TATTGGAATT	1500
TACCGAATGT TTCATGAATT TGGGGTTGTG TATACTAGAC CAGGGCGCAG ACATGACTTT	1560
GTCCAGAGT TACGATTTGA AGATTTTTTA GATAACAGC TATCTATAGA TGAACAGCC	1620



547

AGTTACTATC ATAGGGGAGT ATGTATAGAG GGAGCGGATT CATTGAAAA TATACTAGAT	1680
TTCATTGATT GGCTACCTAA GATTGGGATG AACAGTTTT TCATCCAGTT TGAAAATCCT	1740
TACTCTTTTT TGAAACGTTG GTATGAACAT GAATTTAATC CATATCTAAA TAAAGAACAA	1800
TTTTCAAATG AATTAGTACA AGAATTGAGT GATAGGTTGG ATAAAGAATT GCAAAAAAGA	1860
GGTCTTATTC ATCATCGTGT TGGTCATGGA TGGACAGGTG AAGTTTTAGG TTACTCTTCA	1920
AAATTGGCT GGAATCAGG TCTTAGTATT TCAGAGGAGA AGAAACCCTA TGTCGCTGAA	1980
ATAAACGGGA AACGAGAATT GTTTAATACG GCTCCGATTT TAACCAGCCT GGATTTTCA	2040
AATCCAGATG TAGCTGATAA GATGGTAGAA ATTATCAAGG ATTATGCCAA GAAAAGACCT	2100
GATGTAACT ACTTACATGT ATGGTTGTCG GATGCTCGTA ATAATATTG TGAATGCCAA	2160
AACTGTAGAC AAGAATTGGT TTCGGATCAG TATATTCGTA TTCTCAATCA ATTGGATAGG	2220
GCTTTAACGA GTGAGGGATT AGATACAAAG ATTTGTTTTG TGCTTTATCA TGAGTTGTTA	2280
TGGGCACCTC AGAAAGAAAA ATTAGATAAT CCTGAACGCT TTACCATGAT GTTGCACCG	2340
ATTACAAGAA CATTGAAAT GAGTTATGCA GATGTAGATT TTGACAATTC CATACCTACG	2400
CCTAAACCTT ATATGCGTAA TAAAATTATA CTTCCGAATT CTCTTGAGGA AAATTTATCT	2460
TATCTTTTTG AGTGGCAAAA AGCATTAAAA GGAGATAGTT TCGTATATGA CTATCCTTTA	2520
GGGCGTGCTC ATTATGGCGA TTTAGGCTAT ATGAAAATTA GTCAAACTAT TTACAGAGAT	2580
GTATCTTATC TTTCCAACCT ACATTGAAAC GGGTACATTT CGTGTCAGA ATTACGTGCC	2640
GGATTCCCTC ATAAATTTCC TAATTATGTC ATGGGGGAAA TGCTCTGGAA GAAGACAAGA	2700
AGTTATGAAG AATTGATTGA AGAATACTTT TCTGCTTTGT ATGGGGAAAA TTGGCAGTCT	2760
GTGTGTGAAT ATTTAGAAAA ATTATCCATT TATCTCTCTT GTGATTATTT TAATGCAATT	2820
GGCAGCCGTC AAAGTGATGT TTTAGCGAAT CATTATTATA TAGCTTACAA TCTAGCTGAT	2880
AATTTTTTAC CAATTATTGA GGAAAATATT TCTAAGTTAT TAAATAGTCA AAAGGATGAA	2940
TGGAAACAGC TCAGTTATCA TCGTGAATAT GTTGTTAAGA TGGCGAAGGC TTTATATCTT	3000
CAAGCAACTG GAAAAACAAG GCAAGCTCAA GATGAATGGA GAAATGTGTT GAATTATATC	3060
CGTGGGCACG AATTGCTATT TCAATCTAAT TTGGATGTTT ATCGTGTAAT TGAAGTAGCA	3120
AAAAATTACG CTGGTTTCCA CTTATAAATC ATAAGTATAG AAAATGAACT AAGGTATTCA	3180
GAGAAGATTG ATCCTAAATA TTATGAAATT TAAGGATTTT TAAGATATTT AGGGTCAACT	3240
TTCTATTTAT ATCGTAGCGA AGTCATTTTA ATAATGATGT GTAAAAGATG GATCAAGATT	3300
GAGGAGGAAG AAAGATGAAA TCAAAAGAAG AAATAAATAT GCTTGGTTTT ACAATTGTCG	3360

548

CTTACGCAGG AGATGCAAGG TCAGATTGA TGGATGCTTT GCGCTTTGCG AGAGATGGAT	3420
ATTTTGAACA GGCAAGAGAA TTGGTTGAGT CTGCAAACGA CTCAATAGTG TCTGCCCATC	3480
GAGAACAGAC TAATTTATTA GCGGAGGAGG CATATGGAGA TAATTTTGAA GTGAGCTTTA	3540
TTATGATTCA TGGTCAAGAT ACTTTGATGA CAACGATGCT ATTGTATGAT CAGGTAAAGT	3600
TTTTTATTGA TGAATATGAA CGAATTCGAA AGATTGAAGA ACATATTGGT TTGCAATGAG	3660
GATTAGTCAT GGAAAATTTA CAGGTAAAG CTTACCAGAA GGAGTTTTTA TTAGGAACTG	3720
CTACCGCTGC TTATCAAGTA GAGGGTGCAA CTAGGGTAGA TGGCAAAGGA ATAAATATGT	3780
GGGATGTTTA TTGCAAGAA AATAGTCCGT TCTTACCAGA TCCAGCTAGT GATTTTATT	3840
ATCGTTACGA AGAGGATATA GCTTTGGCGG CAGAACATGG TTTGCAGGCT TTGCGTTTAT	3900
CTATTTCTTG GGTTCGTATA TTTCTGATA TAGATGGGGA TGCTAATGTA TTAGCTGTTT	3960
ATTATTACCA TAGAGTTTTT CAGTCTTGCT TAAAACATAA TGTGATTCCG TTTGTTTCTT	4020
TACATCATTT TGATTGCGCT CAGAAAATGT TAGAAACAGG GGATTGGTTG AACAGAGAGA	4080
ATATTGATCG TTTCATACGA TATGCTCGCT TTTGTTTCCA AGAATTTACA GAAGTCAAGC	4140
ATTGGTTTAC AATCAATGAA CTGATGCTCTC TTGCTGCAGG TCAATATATA GGAGGTCAGT	4200
TTCTTCCAAA TCATCATTTT CAATTATCTG AAGCAATTCA AGCGAATCAT AATATGTTGT	4260
TGGCGCATGC TCTTGCAATC CTCGAATTTT ATCAATTAGG GATTGAGGGA AAGGTAGGTT	4320
GTATTCATGC TTTAAAGCCA GGCTATCCTA TTGATGGGCA AAAAGAAAAAT ATTTTGGCAG	4380
CTAAACGGTA TGATGTTTAT AATAATAAAT TTCTATTAGA TGGAACTTTT TTGGGCTACT	4440
ACAGTGAGGA CACGCTTTTT CACTTGAATC AAATATTGGA AGCTAATAAT TCTAGCTTTA	4500
TTATTGAAGA TGGTGATTTA GAAATTATGA AGAGAGCTGC ACCTCTTAAT ACGATGTTTG	4560
GGATGAATTA TTATCGTTCA GAATTTATTC GTGAATACAA AGGTGAAAAAT AGACAAGAAT	4620
TTAATTCAAC AGGAATAAAA GGACAGTCTT CTTTAAATT AAATGCTCTA GGTGAATTTG	4680
TAAAAAACC TGGTATTCCG ACAACAGATT GGGATTGGAA TATTTATCCT CAAGGGTTAT	4740
TTGATATGTT GCTTCGTATC AAAGAAGAAT ATCCTCAACA TCCGGTCATT TATTTAACTG	4800
AAAATGGTAC AGCCCTTAAA GAAGTTAAGC CAGAGGGCGA GAATGATATT ATTGATGACA	4860
GTAAGAGAAT CCGTTATATT GAGCAACATT TACACAAAGT TTTAGAGGCT CGAGATAGAC	4920
GAGTCAATAT TCAAGGCTAT TTTATATGGT CTTTGCAAGA TCAATTTTCT TGGGCGAATG	4980
GCTACAATAA GCGATATGGT CTTTCTTTTG TTGATTATGA AACACAGAAG AGATATATTA	5040
AGAAAAGTGC TCTTTGGGTA AAAGGGCTAA AACGGAATTA AGGTAGCGA TTTGACTGAT	5100
GTTTAATATG TTTTAAATAT GAGGTTGAAT TTTTATAGG AGGAGTTTTA TGGATAAGCT	5160

549

AGTCGCTGCC ATTGAAAAGC AACAAAGGAA ATTTGAAAAA ATTTCTACTA ATAACATATAT	5220
GATGGCTATT AAAGATGGAT TCATTGCTAC TATGCCTTTA ATTATGTTTT CAAGCTTTTT	5280
GATGATTATT ATTATGATTC CTAAAAATTT CGGAGTAGAG TTACCGAGTC CAGCTATTGT	5340
CTGGATGAGA AAAGTGATA TGTTAACCAT GGGAGTTTTG GGTATTATTG TTTCAAGGAC	5400
TGTTGAAAAG TCATTAGTTG GAAATGTTAA CAGAAAAATG CCTCACGGAA AGGTAATAAA	5460
TGATATTTCT GCAATGTTGG CAGCCATATG TAGTTATCTG GTATTAACCTG TAACGCTTGT	5520
AGTTGATGAG AAGACGGGAT CTACAAGTTT GTCGACAAAC TATTTAGGAT CTCAAGGATT	5580
GATAACTTCG TTTGTCAGTG CCTTTATTAC TGTAATGTT TACCGATTCT GTATTAAGCG	5640
AGACATTACT ATTCATTTAC CTAAGGAAGT TCCTGGGGCT ATATCACAAG CTTTLAGAGA	5700
TATTTCCCT TTTCTTTTG TTTTACTTAT TAGTGGTTTG TTAGATATTG TATCTCGGTT	5760
TAGTTTAGAT GTTCCTTTTG CCCAAGTATT TCAACAATA TTGACTCCTA TTTTAAGGG	5820
GGCAGAATCA TATCCTGCTA TGATGTTGAT TTGGTTTATG TGTGCTTTGC TTTGGTTTGT	5880
TGGAATTCAT GGACCATCTA TTGTCTTACC TGCTGTTACA GCTTTGCAAC TGAGCAATAT	5940
GGAAGAGAAT GCTCAACTTC TTGCAAATGG GCAGTTCCTT TATCATTCTT TAACACCTAA	6000
TTTCGGGAAT TATATCGCTG CTATTGGAGG AACGGGGGCT ACCTTTGTTG TACCATTAT	6060
TTTGATTTTC TTTATGCGGT CTAACAATT AAAATCGGTA GGTAAAGCTA CAATTACTCC	6120
TGTTTTATTT GCGGTAAATG AACCTCTTCT ATTTGGTATG CCTGTTATTT TGAATCCCTA	6180
TCTTTTGTG CCTTTTTC A TGACTCCACC AGTGAATGTA TTTCTAGGAA AGGTCTTTAT	6240
TGATTTCTTT GGAATGAATG GATTTTATAT CCAGTTACCT TGGACCTTTC CTGGTCCCTT	6300
GGGATTGTTA ATTGGAACGA ATTTTCAACT TATCTCCTTT GTATTTTAT CTTTGATTTT	6360
AGTTGTCGAC ATATTGATTT ATTTGCCATT CTGTAGAGCG TATGATAGAC AGTTACTGGT	6420
GAAAGAAGAT ATTGCAAGCT CAAATGATAT TATTTTAGAG GAGGATACAA GTGAAATAAT	6480
TCCTGGTGAG ATAGATGAAA TAAAAAGTAA GGAGTTGAAA GTACTGGTTC TTTGTGCAGG	6540
GTCTGGAACA AGTGCGCAAT TAGCCAATGC AATTAACGAG GGGGCTAACT TAACAGAGGT	6600
TAGAGTGATT GCGAATTCAG GACCGTACGG AGCTCATTAT GATATTATGG GTGTTTATGA	6660
TTTAATTATT CTGGCCCCAC AAGTTCGGAG TTATTATAGA GAGATGAAGG TGGATGCAGA	6720
AAGATTAGGT ATTCAGATAG TTGCTACCAG AGGAATGGAA TATATTCATT TAACAAAGAG	6780
TCCAAGTAAA GCCTTACAAT TTGTATTGGA GCATTACCAA GCTGTGTAGT AAGTTTTTCC	6840
ATCTTTTATT TGAGTAAAGA TTTTGTTTAC AGATAGGCTT GGATTTAAAA ACGTTCCCCC	6900

550

TTTTTTAATA TAAGAATCCC TCTTTCACAA TTGTAAAAAG AGGGATTTTG TATTTTATCT	6960
CTTAGACCAA GTTCTCTTCA TAAAGAGAAG GAGGATTGGG TAAATCTCCA AGCGCCCTGC	7020
AATCATTGCA AAGGATAGGA GAATTTTGA GATGGGACTA AAGATTGAGA AACTAGAAGT	7080
GGTTCCTAGA ATAGGCCCGA TATTATTGAA ACAGCTAAAG ACAGCGCTGG TCACGACCAG	7140
AAAATCATTG CTATCTAGGC TGACAATAAA GATAAGCGCT AGCAAAATCA TAGCATAGAT	7200
GACAAAGTAC TTGAGAATCT TATGCTGGGT ATCTTTGTCA ATCACCCTTT TATTAACATG	7260
GAGGGTCAAA ACACGGTGGG GCGATAGGAT TGACAAAATT TGGT'TTTTG CAAT'TTTGA	7320
AAGGATGAGG CCTCGAATAA TCTTGAGTCC ACCTGCAGTT GATCCAGCAG AGCCACCGAT	7380
TGCCATGAGG AAAAGGAGGA TAAACTGGGA GAAGAGGGGC CAGT'TGGTAA TATCTCCATA	7440
TCCAAAACCA GTTGTTGTAA TGATGT'TGGA AACCTGGAAG AAGGTCATT' CAAAGCTCTT	7500
TGAAAACCTT GGGTAGAGGT AGAGGGTGTT GAGGCTAATC AAGCCTGTAG AAACCAGTAC	7560
AATGACCAAG TAAGCCCTAA GCTCTTCATC TCCAAAGAAG GCCT'TGATGC GACGGAGCAT	7620
GAGGTAGTAG TAGAGGTGA AATTTACTCC AAAAACCAGA ACTCCGATAC TGACCAGATA	7680
GGTAATCAGT GAGCTGCCAT AGTGGGCAAT TCCGTCTGTA TAGACGGTAA AGCCTCCAGT	7740
TCCCGCTGTC CCCATAGCAA TAACAAAAC' ATCGTAGAGA GGCATACCGG CTAGATAATA	7800
GATGATGACA AAGAGGGAGA AGAGAGCTAG ATAAAGGAGA TAGAGAATCT GGGCAGTGTT	7860
TTT'TAGTTG GATACAACCT TGCCAAAAAC AGGACCTGGA ACCTCAGCCT TCATCACCTC	7920
TAGGTGGCTA TTTTGGCAT TGTCCATAAT AGCAAGTGCA AAAACAAGCA CTCCCATCCC	7980
TCCAATCAAG TGGGTAAAAC TTCGCCAGAA GAGGAGGGAA CGGCTGAGAA CCGAAACGTC	8040
GTTCAAAATA CTGCTCCAG TAGTTGTAAA TCCAGAACTA ATTTCAAAAA AGGCATCAAT	8100
AAGGCTGGGG A'TTGCCCGAG AAAAGACAAA GGGGAGACCA CCAAAGAAAG ACCAAAGGAT	8160
CCAACAGAGG GCAACGATCA AGACTCCCTC CTTGGCATAA ATCCGTGTAT TTTTGGCTT	8220
CTGTAAACTC CCTGAACCGC CTAACAATAC GAGAATCCCT ATGGTCGAAA AGAGGGCTGT	8280
AAAGACTTGG CTCGATTAC GGTAAATAGAC AGCAATCGCA ACAGGAACCA AAAGAAGAAC	8340
AGCTTCAATC AAAAGTAATT TTGAAAGGAG GTAACGAATC ATACTTTTAT TCATTTCTTA	8400
CCTCGCGATC AAGTCATAAA TCTTGGTGAT GTTTGGCAAC AAGGTTGTTA CTAGGAGCTT	8460
GTCTCCAACT TCCAACATAT CCTCCCCAGT TGGGAAAATA GTCTTGCCCT TTCGAATAAT	8520
GGCTGCAATA AGAACCCCTT TTTTCAATTT CAGTTGAGAA AGAGGTTTGG CAGTCATTTT	8580
ATTGGCTTCC TTGATATGGA ATTGCAGGGT TTCCATTTGG CCATTGGCTA GATGGTGCAT	8640
AGCTTGAAGG TCTGAATACT GGGCATTAAC TCGACCACGA ATAAAGTGCA TAATCGTATC	8700

551

TACAGCGATG CTTTtaggtg TGATGATACT TGAAAAATCA GGCGCATTGA TAATCTCGAG	8760
GAGACTGGTA CGATTGACCT TAGTAATATT TTTCTGTACA CCTACCCCTGT CAAGGAACAT	8820
AGATGTAATC AGATTTTCCT CATCGACTCC TGTTAGAGTC GCAACGGCAT CATAGTGTTC	8880
AGCACTTTCT TCCAGCAGGA TATCTTTTGC GGTTCATCT CCTTGAACGA TGTAGAGATT	8940
TGGGAATTTT TCGCTAAAGA AGCTGGCGAT TTCAGGATTG ATTTCAATGA CTTTGTATC	9000
GATACGACTA TCTTGAGAA TACCAAGTAG ATAATAGGCA ATTCTACCTG CCCCAACGAT	9060
GAGAAGGCTC TTCACGGCGC GTGATTTAAA ATAATTATGG AAGAGTATCA TATCGACACG	9120
GTTACCAGTG ACAAAGATTC TATCTTTATC CTGTACAGTC ATGTCACCGC TTGGAATGAT	9180
AATTTGATGA TCCCTCTCTA TCGCACAGAC AATGACATTA CCAAATTTTT TACGAAAATC	9240
AGAAATGGGC ATTTGGCAAA GACCGCTGGT GGACTTGACG ACAAATTCCA TGAGGCTAAC	9300
GCGTCCACCA GCAAAGCGTT CGACAGACAG GCGTTGGGG AAGTCAATGA TATCGCGAT	9360
AGCGCGGGCA GCCAAGAGCT CAGGATTAAC GATAAGAGAA AAACCGAGAA TATCTTTTC	9420
CTTGAAATAA GAGTTAGAAT ATTCAGGGTT CCGCACCCGA ACGATAGTTT CTTAGCTCC	9480
CATTTCTTGT GCTAGAACTG CTGCAATCAT GTTGACTTCA TCGTGCTCAG TCAGGGCGAT	9540
AAAGATATCA CAATCTTGA CGCTGGCTTG CTCAAGAATG GCAAAATCGG CCCCGTTACC	9600
AAGGATACCA ATGATATCAA AGCGACTGAC AATATGATTG AGAACAGCTT CGTCTTGCTC	9660
AATCAGCAAA ACATCATGCT TTTCTGCAAC CAAGGAGCGA CAGAGGGCAA AACCAACTTT	9720
TCCCCCTCCG ACAAGGATAA TTTTCATAAT AAAACCTACT TTTTCATGAT GTAACATCA	9780
TACCCTTTTT CAAGAAAAAA TGCACCTACT AGCTAATAAC AAGAGTTTTT AGTGAAAATT	9840
CGCTATAAGG TAAAACTATA CCCTAACCAA TTGAAATAGC TATTAGCGAC TTTCTCTGAA	9900
ATATGGTATG ATAAAGGATA TACAAGGAGA TAAATGAAT AATAATTTAC TGGTATTACA	9960
ATCAGACTTT GGTCTGGTTG ATGGTGCGGT ATCGGCTATG ATTGGAGTGG CTTTAGAAGA	10020
GTCTCCAACC TTAAAAATAC ATCACTTGAC GCACGATATC ACGCCTTATA ATATTTTGA	10080
GGGGAGCTAT CGTCTCTTTC AGACGGTGGA TTAAGGCTT GAGGGAACGA CGTTTGTATC	10140
GTTTGTGCGT CCAGGTGTCG GTTCGAAACG TAAGAGTGTA GTTGCCAAGA CTGCAAAAAA	10200
TCAATACATT GTCACGCCAG ATAATGGGAC GCTTTCCTTT ATCAAGAAAC ACGTTGGCAT	10260
TGTAGCCATT CGTGAGATTT CTGAGGTGGC CAATAGGCGT CAAAACACAG AGCATTCTTA	10320
TACCTTCCAC GGTGCGTATG TCTATGCCTA TACTGGTGCT AAAGTGCCA GTGGTCACAT	10380
TACTTTTGAG GAAGTAGGGC CAGAGCTCAG TGTGGAACAG ATTGTAGAGC TTCCAGTCGT	10440

552

AGCGACCATC ATAGAAGATC ATCTGGTGAA GGGAGCCATT GATATTCTGG ATGTGCGTTT	10500
CGGTTTCGCTT TGGACCTCTA TCACACGGGA AGAATTTTAC AAGCTGGAAC CAGAATTTGG	10560
TGATCGTTTT GAAGTGACCA TCTATCATGC TGATATGCTG GTCTATCAAA ATCAGGTTGT	10620
CTATGGCAAA TCATTTGCAG ATGTGAGAAAT TGGGCAACCs ATcTTTACrc TCAGCaTCTc	10680
CGATTAGCTG GGCAATTCGT TCTAGTTGGA TTTCGTCAAT CAAGGT	10726

(2) INFORMATION FOR SEQ ID NO: 67:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7163 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTATCTTTAA CGATATCAAT CAAGATCTGG TCAATAAAGG GATTGGGGCT TATCGTGAAG	60
TTGGCATCCA AGCCCATGGA TATGTCTGTG ACGTGACAGA CGAGGACGGT ATCCAAGCCA	120
TGGTCAAGCA AATCGAACAA GAGGTTGGTG TCATTGACAT CCTCGTTAAT AACGCTGGTA	180
TTATCCGCCG AGTTCCAATG TGCGAAATGA GCGCCGCTGA TTTCCGTAAG GTCATCGATA	240
TTGACTTAAA CGCACCATT ATCGTTTCAA AGGCAGTTAT TCCTTCTATG ATAAAGAAAG	300
GGCATGGAAA GATTATCAAT ATTTGTTCTGA TGATGAGCGA ACTGGGACGT GAAACAGTTA	360
GCGCTTATGC TGCTGCTAAA GGGGGCTTGA AAATGTTGAC CCGCAACATT GCGTCTGAAT	420
ACGGTGGAGC CAATATCCAA TGTAACGGAA TTGGACCGGG TTATATTGCC ACTCCTCAAA	480
CAGCACCTCT TCGTGAATTG CAAGAAGATG GTTCTCGCCA CCCATTTGAC CAGTTTCATCA	540
TTGCAAAAAC ACCTGCTGCA CGTTGGGGAA AFACTGAAGA TTTGATGGGC CCTGCTGTCT	600
TTCTCGCTAG TGATGCCAGC AATTTTGTCA ATGGCCACAT CCTATATGTA GATGGCGGTA	660
TCTTAGCCTA CATCGGAAAA CAACCTGAGT AAAAATAGAA AGAAGATCTT ATGAAAATCG	720
CATTAATCAA TGAAAATAGT CAAGCTAGCA AGAATCACAT TATTTACGAT AGTCTAAAAG	780
AAGCGACAGA TAAAAAAGGC TACCAATTAT TTAACATGG TATGCCGTGA GAAGAAGGAG	840
AAAGTCAATT AACTTATGTG CAGAACGGAC TAATGGCTGC CATCCTTTTA AATACAAAGG	900
CAGTTGACTT TGTTGTTACC GGCTGTGGTA CGGCTGTAGG GGCTATGCTT GCTTTAAACA	960
GCTTCCCTGG TGTTGTCTGT GGTCTAGCAG TGGACCCAAC TGACGCTTAC CTTTATTCTC	1020
AAATCAATGG TGTAACGCC TTGTCTATCC CTTATGCCAA AGGATTTGGC TGGGGGGCAG	1080
AACTGACCCT CAAATTGATG TTTGAACGCT TATTTCCTGA AGAAATGGGC GGTGGCTACC	1140

553

CAAGAGAACG TGTAATCCCT GAACAACGCA ACGCTCGTAT CTTAAACGAG GTGAAACAAA	1200
TCACCCACAA TGATTTGATG ACCATCCTTA AAATAATCGA CCAAGACTTC CTCAAAGACA	1260
CCATCTCTGG CAAATACTTC CAAGAATACT TCTTTGAAAA CTGCCAAGAT GATGAAGTTG	1320
CTGCTTATTT GAAAGAAGTA TTAGCCAAGT AAAGCTATTC TAAACCAGAA AGGAACTAAT	1380
GGATGACGAA AATATTACTG TTTGGCGAAC CATTAAATTCG AATTCACCA TTAGATGCCA	1440
CCAGTATCGG CGATCATGTT GCCAGTTCGA CTTATTTTGG CGGATCAGAA ATTAACATCG	1500
CTTGTAATTT GCAAGCCCTG GGTATCTCAA CGAAAGTTTT TACCGCACTC CCTGCCAACG	1560
AGATTGGAGA TCGTTTTCTC ACATTCTTGA AACAGCACCA AATCGATACC AGTTCAATCT	1620
GTGGGCTTGG CGATCGAATC GGCCTCTACT ATTTGGAGAA CGGCTTTGGT TGTCGTCAAA	1680
GTGAAGTTT CTACGATCGT AAGCATACGA GTATCAGCCA GATTCGGCCA AACATGCTAG	1740
ATATGGATTC TCTCTTCAG GGGATTAGCC ATTTTCATTT TAGTGGAATC ACCGTAGCTA	1800
TCCGTCAAGA GGTCCGTGCG ATCCTTCTCC TACTCTTGA AGAAGCCAAG CGCCGAGGAA	1860
TTGTCGTTTC AATGGATCTC AATCTGAGAA CAAAGATGAT TTCAGTCCTA GAAGCCAAGT	1920
ATGAATTTTC TAAGTTTGCA CGTTTTACTG ACTATTGCTT CGGTATTGAT CCTCTCATGA	1980
TTGATGACCA AAATCTAGAG ATGTTTCCAA GAGACAGTGC TAGCCTAGAA GAGGTGGAAA	2040
ATCGCATGCG ACTTTTAAAA GAAGCCTATG GTTTCAAGGC CATTTTCCAT ACCCTCCGCT	2100
CTAGTGATGA GCAAGACAAA AATGTCTATC AAGCCTATGC TCTAGAAGAA CTATTTGAAG	2160
AGTCTGTCCA ACTAAAACT GCAGTCTATC AACGAATTGG TAGCGGGGAT GCCTTTATAT	2220
CTGGTGCCCT TTACCAACTA CTCCATCATT CCTCCCTAAA AACTACCATT GACTTTGCAG	2280
TTGCGAGCCG AACTCTCAA TGCACCTTC CAGGAGACCA TCTCTCCACT TCCTCAACTA	2340
GTATTGAAAA TTTACTGGCA AATGCACAAG ATATCATTCG TTAGGAGAAT TACATGACCA	2400
AATCAGATAC GATTATTGAA CTAAAAAAC AAAAAATTGT CGCTGTTATT CGAGGAAATA	2460
CAAAGGAAGA AGGACTACAA GCCTCGATTG CTTGTATCAA GGGCGGTATC AAAGCTATTG	2520
AAATCGCCTA TACCAATCAG TATGCAGGAC AAATCATCAA GGAACCTGTA GACTTGTATC	2580
AGGACGATCA GAGTGTGTTG ATCGGTGCAG GTACTGTGCT TGATGCCGTA ACTGCTAGAG	2640
ATGCCATTCT AGCTGGAGCA AATTACGTTG TTTCTCCATC TTTCCATGCT GAAACTGCGA	2700
AAATGTGCAA TCTCTACAGC ACACCGTACA TTCCAGGCTG TATTACCCTC ACAGAGATCA	2760
CGACTGCACT TCAAGCCGGT AGTGAAATCA TCAAACCTTT CCCAGGTAGT ACTCTCAGTC	2820
CAGCATATAT CTCTGCAGTC AAGGCACCGA TCCACAAAGT TTCCGTAATG GTAACCGGAG	2880

554

GAGTCGGCCT AAACAACATC CCTCAATGGT TCGCTGCTGG TGCAGATGCC GTTGAATTG	2940
GTGGCGAACT CAATAAACTC GCTTCCCAAG GCAACTTTGA CCGCATCAGC GAGATTGCCC	3000
AACAGTATAT TACACTCAGA TAAAATCATA ACTACCCGTC TAACGGGTGG TTTATCTCAG	3060
AGCTATAAGC CCAAATCATC AGCCAGCGCC TAAAGACGCT GGCTTTCACG TTGTTCAAGC	3120
CTTATTGCTC TTGACTCGTC ACTTGCTCT TTAAGAGACT TTGGTATTAC TTACCACTAT	3180
CCCTAAAGGG ATCCTCATAT TCTTTTACAC TCAATTTATC TAGTGCTATA GTAGATTGAA	3240
ACTGGAATAG TACACCTCTG CTTCTAAAAC ATTGTTAAAA ATCGATTGTA CTGTCCTGAT	3300
CGATTTTGTC CTGTTCTTAT TTCATTTTAC TATATATCAT ACTTTACTCG TTCTCAAATT	3360
TTCATACTCA TGAAGAAATC ATCCACTCGA TAATTTCTTT AATCTTGACT ATATTTCTTA	3420
ATTGTGGCTT CATTAAAGCC TACTGGACTT ACATAATAAC CTTCTCCCA GAAATGCCGA	3480
TTCCCAAAC TGTACTTGAG ATTGGCGTGT TTGTCAAACA TCATGAGTGC ACTTTTGCCT	3540
TTTAAATACC CCATAAACT TGAAACACTT AGCCTCGACG GAATACTGAC TAACATGTGT	3600
ACATGGTCTG GCATTAAGTG ACCCTCGATC ATTTCAACAC CTTTATAACT ACACAAGCGA	3660
TGAAATATTT CGTCTAACT ACTTCTATAT TGATTATAGA TGACTTTTCG TCTATACTTA	3720
GGGGTGAACA CAATATGATA GAACACCTCC ACTTTGTGTA TGATAAACTA TGAGTCTTTT	3780
GTGCCATATT TTTTCTCCTT TCGCTTTACA ATTGGATTGA ACACCTTTAT TGTATCGCGT	3840
TTGGAGTTTT TTTGGTATAA CCTTCGACGC GCACCCGTAT AGCGGGTGGT TGTTTTGTCT	3900
CGCACCTCAC GGAGCGAGAC GGACTAATAT AGTGGAGTGA AATAGGATAC GAACAAATTG	3960
ATTAGGAAAA TCAAATGAAT TTATAGAAAT CTTTTCAGCAG TTATAACGTT CTATTCTAGT	4020
TTCAAAACGC TATAGTCACA TAATAATGAA GTAAAAAAGG ATAAGTATCA ACTTATCCTT	4080
TTTTAAAAGA AAAATCCGAA GATATTTGGC CTTCTTCGGA TTTTCTCTAT TTTCCACAGT	4140
TTCATGTAAT TCATCTAGAT GATGAACAAA TTAGTTGTTT TTTCCTCTAC GGAATAGATA	4200
AAATGCCCCA AGTAGCAAGA ACCCTAGACT TGCCAAGATT GACTGACCTT CTCCTGTCTG	4260
AGGGAGATTC TTTTGATCCG AATGGTTCTT TTCCTCTTCA GATTTTTCCT TTTCTTTTGA	4320
ATTCTGTACT TGTGGCTGAG CTGCTTGCTC TAGCTTTTTA AAGACTTCCT GATCTGGAGC	4380
TGATTCCTGG GTTTCAGGAT TATAGTAGGC AATCTTATAT TCATCCCTT CTTTTCGAAT	4440
GGTATAGACT CCACGTTTCA AAACCTTGAA TTGGTTGGAA ATAGTAGAGA CAGAATCATC	4500
ATATTTTACA ATGCCCCAAA CTCCTTGTTT AGCATCATAA ACAGACTGAA GGGTTTCGTT	4560
ATTTTCGATG AGGCTACTTT CTAACCTTTT TATCATTTGA TTGAAGGTGG CACGATCCAC	4620
GTTAGGAATG AGCATATAGC CATAAGAAATC TCTATTTTGC TTATGAGCCT GACTAATCGT	4680



555

AAGAAATTCA TTTTCAACTT CCTTGTCTGA CTGTCCTTCA TTGATATCCT TCCAGGCTCC	4740
CTTTTGCAAA GCCTTACTCA TACTGATTGA ACTCTTCTTA AAGAAAAAGT AACCAATATT	4800
CTTTTTCGAA TCGAACGATT CTA AAAAGAC ACTTTGGGT TCAGGATAAT CCTTTTCTTG	4860
TTCTGTAAGG GAGGCTTCTT TATCATTGAC ATAGACTTTA TATGGATTAC CTGATTCCAG	4920
TTTTCTCTGG TCAATTGTAG TTGCAGCAGT ATCTGTTGAA GTGTTTTGGA TATTGCTTCC	4980
TAAAAAGGCG ATCTTATCCT TTAGCATAAA CCAGCTCTTA TGAGCAGTCA ATGTTTGATT	5040
CCAGTTGGTG AAATCCATGG TTGCTGTCCG ATTGGCATCA TCTAGTTTGC TCGTTCCAAC	5100
GAAAGCAGAC GGTAAAACTT TACCTGTATC GCTATCCGCT CTCTTAGCAT CCGTCTCTGT	5160
TGTACCAGGC ATCTTATATG GATTAACTGT TGGCCAGTAG CCATCGCTAT AGTGACTCAA	5220
ATCGCCATTG TAAAGATAGA ACATCCCATC ACTCGTATAC CAACCACGTT TATTTTCCTT	5280
GTTTATGTGT TCGTAATTCA AGGTACGACT GGAAAAGAGT GACAAGCCAA ATCCAAAACC	5340
TTTCTCTGCA TTGTACATGG CTGTTTTATC CATCTTGTTA AAGGCAGATA GGTAAC TTGG	5400
TCTTGGAACA CTTGCGACTC CTGCATCACT TAACAAGGAT TGCATCAAAC TGATATCCTT	5460
ATAAGTCTTC AAATTCTTAA AGACATCATA ATA ACTATCC GATTGAACAA TGGTCTTCAC	5520
AAGACTCTGC AAACATTGTT TGGTTTCTCC TTCAGACATA TCCGCTATTC GGTGAATCCC	5580
TCTTAGTACT TCTACTGCGG CCACGTGCCC CTCGCTATTT GCACGACTGA TCGAGCGTCC	5640
ACGACTCATA TCCATCAACT CTCCATTAC CAGCAAAGGA GCAAACGATT TATCAATCCA	5700
GTGGTACATG GTTTGCATTT TATCTTTATC GATTGGATTG TTGGTCTTTT GAATGACTGG	5760
CAACAGTTGA GACAGGCCAT CAATCAAAC ATTCCATAA GCACCCGTAT AGGCAACATT	5820
GGTGTGTCG ATATAGGATC CATCTTGATA AAAACCTTCA CCTTGGTCTA CCAACTTGAA	5880
CACTTGCTCA ATCGAGCGAA TGGTAGAAGA AATTTCTTGA TCATCCTTAC GCAGTAAACC	5940
AGCTATTACT TTTACCCTTC CCATATCAAC TAAGTTTCCA CCTAGAGCCT TGAATGGGTT	6000
ATCAGTCGTC TTTCCGAAAT GTTCGGGATC TGGTACAAAT TTTTCAATCA CATCTGTATA	6060
TTTTTTAATT TCCTCATCAG AGAAGTATTC TTTTATCAGA GACAAGGTAT TGTGATGGC	6120
ACGAGGTGTA CCGATTTCAT AATCCCACCA GTTCCCAACA ATGCTCTTTT CACTATTGTA	6180
GACATGTTTA TGCATCCATT CCATGGAATC CCTGACTGTT CGAACGACAG TTTTATCTTG	6240
ATAATAACGA GAAGAAGGAT TGGTCACTTG CTGGCCATC TCCTCCAATT TCCGATAAGT	6300
GGCAGTCAGA TTTGCAGACC TTTTATAATT TGAAAATTTT TCCCACAAAT AGGTGCGGTC	6360
CGCCTGACTT GAAATACTGG ATAGGCTATC AGTACCTTT CCTTCCAATT CCTGGTTTAA	6420

556

TTTGGCCATC TGTTCAATTT TAGAATCATA GTATTGATTC CCAGCGATGA TGCCATTCCA	6480
GTCATCCAAA CGGTCGTGT ATGCATCCTT AACAGAGGCC AGAATCTTCA AAGGAATCTT	5540
TTTCACTTCC TTGCCATCTT TACTGACAAT GACATTGGTT GTCCCTTCCT TAAGAGGTTT	6600
TAAAATTCCA TTTTGTACTG AAGCAACGTC AGGATTTTCT ACCTTATAAG TATAGTCCGC	6660
AAGAGAAAAA ACATGTTTTT TTCCAATTGG TAAATCAATC TTTTCCTCAA GCTGTTTATC	6720
TGTTTGAGAA TCCTCAGAAA GCTGGTCTGC TACCTCTACC AGCTCAATAT CCTTAAAGGA	6780
AACAGTCCCA GTTCTGTTT CATAGAATAA CTCCAGCTTG ATTTTATCAA CATCTAAAGT	6840
CGGGCTATAG TCTGCTTCAA TGGTCTGCCA GTCCTTTGTT CCTGACGTCG TTGCAGAATT	6900
CCACAATCGC TTGTCTTAC CACTTTCCTC AATGATACGA ACTTTGGCAA TCCCATTCTT	6960
ATTATCTGTT TTAATCTTGA AACGCAGTTT ATACTTTTTC TTAGCTTCAA TAGGAACCAT	7020
ACGGTGAAGC GCTGCCCTTA ATTTCTCATG GCTTGAGATA GTGATAGCCC CATCCTTAGC	7080
CTCAATGACT CGAGTTGAGG CATCTGCACT ATTCTTCTGG TCTACCCAAG CTGACCACCC	7140
CCTGAGCTTT GCTTCTGTC CGG	7163

## (2) INFORMATION FOR SEQ ID NO: 68:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9244 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

CGTTATAACA TACATGTAAG CGGTACCCAA AATGGTGCCA AGTCAAAATT TTTAAGGAGG	60
AAAAATACATG TCTTCACATC CAATTCAGGT CTCTCAGAA ATTGGGAAAC TGAAAAAAGT	120
TATGTTGCAC CGTCCAGGCA AGGAGTTAGA AAACCTGTTG CCGGACTATC TTGAAAGGCT	180
TCTTTTGGAT GATATTCCTT TCTTGAAGA TGCTCAAAAA GAACATGATG CATTTGCCCA	240
AGCTCTTCGC GATGAAGGAA TTGAGGTTCT CTACCTAGAA CAACTCGCTG CTGAATCATT	300
GACCTCTCCA GAAATCCGCG ATCAATTTAT CGAGGAATAC TTAGACGAAG CCAACATCCG	360
TGATCGTCAA ACCAAGGTTG CTATTCGTGA ATTGCTTCAC GGCATCAAGG ACAACCAAGA	420
ATTGGTTGAA AAAACAATGG CTGGGATTCA AAAAGTTGAA TTGCCAGAAA TTCCTGACGA	480
AGCTAAAGAT CTAACGACT TAGTTGAATC AGAGTATCCA TTTGCAATTG ACCCGATGCC	540
AAACCTCTAT TCACTCGCG ACCCATTTGC AACAATTGGA AACGCCGTAT CGCTTAACCA	600
CATGTTTGCA GACACTCGTA ACCGTGAAAC ACTCTACGGT AAGTATATCT TCAAATACCA	660

557

CCCAATCTAT GCGGAAAAAG TGGATTGTTGGT CTACAACCGT GAAGAAGATA CGCGTATCGA	720
AGGTGGAGAC GAGTTAGTTC TTTCTAAAGA CGTCTTGCA GTAGGTATCT CTCAACGTAC	780
AGACGCAGCT TCTATCGAAA AACTTTTGGT CAACATCTTC AAGAAAAATG TTGGCTTCAA	840
GAAAGTTTGG GCCTTTGAAT TTGCTAACAA CCGTAAATTC ATGCACTTGG ATACTGTCTT	900
CACTATGGTA GACTATGACA AGTTCACTAT TCACCCAGAA ATCGAAGGCG ACCTTCACGT	960
TTACTCAGTT ACTTACGAAA ACGAAAAACT TAAATCGTT GAAGAGAAAG GTGACTTAGC	1020
TGAACTTCTT GCTCAAAACC TTGGTGTAGA AAAAGTTCAT TTGATTCTGT GCGGTGGTGG	1080
CAATATCGTA GCAGCTGCGC GTGAACAATG GAACGACGGT TCTAACACTT TGACCATCGC	1140
ACCTGGTGTG GTAGTTGTTT ATGACCGCAA TACCGTGACC AATAAGATTT TGGAGAATA	1200
CGGGCTTCGC TTGATTAAAG TTCGCGGAAG TGAATTGGTT CGGGGCCGTG GTGGACCTCG	1260
TTGTATGTCT ATGCCATTTG AACGTGAAGA AGTGTAAATG CTGTTTCGATA TTCGTCAATA	1320
GAAAAATGTA AAAATAGAAA GAGGAAATAA TAAATGACA AATTCAGTAT TCCAAGGACG	1380
CAGCTTCTTA GCAGAAAAAG ACTTTACCCG TGCAGAGTTA GAATACCTTA TTGGTCTTTC	1440
AGCTCACTTG AAAGATTGA AAAAACGCAA TATTCAACAC CACTACCTTG CTGGCAAGAA	1500
TATCGCTCTC CTATTTGAAA AAACATCTAC TCGTACTCGT GCAGCCTTTA CAACTGCGGC	1560
TATCGACCTT GGTGCTCACC CAGAATACCT CGGAGCAAAT GATATTCAGT TGGGTAAAAA	1620
AGAATCTACT GAAGATACTG CTAAGTATT GGGACGTATG TTTGACGGGA TTGAATTCGG	1680
CGGATTCAGC CAACGTATGC TTGAAGAATT GGCAGAATC TCAGGCGTTC CAGTATGGAA	1740
CGGTCTAACT GACGAATGCC ACCCAACTCA AATGCTCGCT GACTACTTGA CTGTTCAAGA	1800
AAACTTCGGT CGCTTCGAAG CCTTGACATT GGTATACTGT GGTGATGGAC GTAACAACGT	1860
TGCCAACAGC TTGCTCGTAA CAGGTGCTAT CCTGGTGTC AATGTTTACA TCTTCTCACC	1920
AAAAGAACTC TTCCAGAAA AAGAAATCGT TGAATTGGCA GAAGGATTTG CTAAAGAAAG	1980
TGGCCACAT GTTCTCATCA CTGAAGATGC TGATGAAGCA GTTAAAGATG CAGACGTTCT	2040
TTACACAGAC GTTTGGGTAT CAATGGGTGA AGAAGACAAA TTCGCAGAAC GTGTAGCTCT	2100
TCTTAAACCT TACCAAGTCA ATATGGACTT AGTTAAAAAA GCAGGCAATG AAAACTTGAT	2160
CTTCCTACAC TGCTTGCCAG CATTCACGA TACTCACACT GTTATGGTA AAGACGTTGC	2220
TGAAAAATTT GGTGTAGAAG AAATGGAAGT AACAGACGAA GTCTTCCGCA GCAAGTACGC	2280
TCGCCACTTC GATCAAGCAG AAAACCGTAT GCACACTATC AAAGCTGTTA TGGCTGCTAC	2340
ACTTGGAAC CTTTATATTC CTAAAGTATA ATTTAGATA ATAAACCGTC TACCAACAGC	2400

558

TATGAGGGCT GCGACTAATA GCTTTAGTCC GGTCTCTTT TATGTAATGG TAATCTATTA	2460
TTTCTTATAA AATATGTGAA AAATCATTA AATTGAAATCT AAACGCATTC TATTGAGTGT	2520
GATAAAGGAG AATTTATGGC AAATCGTAAA ATTGTAGTAG CTTTGGGAGG AAATGCGATT	2580
CTTTCTTCTG ACCCATCAGC AAAGGCTCAA CAAGAAGCTT TAGTTGAAAC AGCTAAGCAT	2640
CTTGTAATAA TGATTAAAAA TGGAGATGAT CTGATTATCA CTCACGGTAA TGGACCTCAA	2700
GTTGGGAATC TCTTGCTCCA ACATTGCGCA TCAGACTCTG AAAAGAACCC TGCCTTCCCA	2760
CTCGACTCAC TTGTCGCTAT GACAGAAGGT AGCATCGGTT TCTGGTTGAA AAATGCTTTG	2820
CAAAATGCTC TCTTGATGA AGGCATCGAA AAAATGTTG CCTCTGTTGT AACGCAAGTT	2880
GTCGTAGATA AAAATGATCC AGCTTTTGT AACTTGAGTA AACCAATCGG TCCTTTCTAT	2940
TCAGAAGAAG AAGCAAAAGC AGAAGCCGAA AAAAGCGGAG CGACTTTCAA GGAAGATGCT	3000
GGCCGTGGCT GCGTAAGGT CGTTGCCTCA CCAAACCTG TTGACATCAA AGAAATTGAA	3060
ACCATCCGTA CTCTTTTAAA TAATGGTCAA GTCGTCGTAG CTGCAGGTGG TGGCGGTATT	3120
CCCGTCGTCA AAGAAAACAA TGGACATTTG ACTGGTGTCTG AAGCGGTTAT TGATAAAGAC	3180
TTGCGTTCCC AACGTTTGGC AGAATTGGTT GATGCAGACC TCTTCATCGT TTTGACAGGT	3240
GTAGATTATG TATTTGTAA CTACAACAAG CCAAACCAGG AAAAATTGGA ACATGTGAAT	3300
GTTGCCCAGC TGGAAGAATA TATCAAACAA GATCAGTTTG CACCAGGTAG CATGCTTCCA	3360
AAAGTAGAAG CAGCTATCGC TTTTGTCAAT GGTCTCCAG AAGGAAAAGC AGTTATTACT	3420
TCCCTTGAAA ATCTAGGCGC CTTGATTGAA TCTGAAAGCG GAACAATTAT TGA AAAAGGA	3480
TAAGTTGTTT TACTAATAAG ATGTATTCTA TTTCTAGTAT CTTTATATCA AATTAGAAAT	3540
TATTCTTGAA AACATGTACA ATATTTCAAA AGATACTAGT TTTAGACTTT AATATGGTAA	3600
AACAAATATA AATAGAAAGC GTTTCTTGA ATGTTTATTT AAGAAAGTAG TTGTTTTTTT	3660
ACACTTTGTT AGACATCAGG AGGAAAAACA AATGAGTGAA AAAGCTAAAA AAGGGTTTAA	3720
GATGCCTTCA TCTTACACCG TATTATTGAT AATCATTGCT ATTATGGCAG TGCTAACTTG	3780
GTTTATCCCT GCGGGGGCCT TTATAGAAGG TATTTACGAG ACTCAGCCTC AAAATCCACA	3840
AGGGATTGTTG GATGTCCTCA TGGCACCGAT TCGGGCTATG CTAGGTACTC ATCCAGAGGA	3900
AGGTTGCTC ATTAAGAAA CGAGCGCAGC GATTGATGTA GCCTTCTTCA TCCTTATGGT	3960
TGTTGGTTTC CTTGGCATTG TCAACAAAAC TGGTGCTCTT GACGTAGGGA TTGCCTCTAT	4020
CGTGAAGAAG TATAAGGGCC GCGAAAAAAT GTTAATTTTG GACTGATGC CTTTGTGTC	4080
CCTCGGTGGT ACAACTTATG GTATGGGTGA AGAAACAATG GCCTTCTATC CACTCCTTGT	4140
GCCAGTTATG ATGGCCGTTG GTTTTGATAG CCTGACTGGT GTTGCAATTA TTTTGCTCGG	4200

559

TTCTCAAATC GGCTGTTTGG CATCTACTCT GAATCCATTT GCGACAGGTA TTGCTTCAGC	4260
GACTGCGGGA GTTGGTACAG GGGACGGTAT CGTACTTCGT CTGATCTTCT GGGTTACCTT	4320
GACTGCTCTT AGTACTTGGT TTGTTTACCG TTATGCCGAT AAGATTCAAA AAGATCCGAC	4380
TAAGTCACTG GTTTATAGTA CTCGCAAAGA AGATTTGAAA CACTTTAACG TAGAAGAATC	4440
TTCATCTGTA GAATCTACAC TTAGCAGCAA AAAAAATCA GTTCTCTTCT TATTTGTGTT	4500
<hr/>	
GACATTCATC TTGATGGTAT TGAGCTTCAT TCCATGGACA GACCTTGGCG TTACCATTTT	4560
TGATGACTTT AATACTTGGT TGACTGGTCT TCCAGTTATT GGTAATATTG TCGGTTTCATC	4620
TACTTCTGCA CTAGGTACTT GGTACTTCCC AGAAGGCGCA ATGCTCTTTG CCTTTATGGG	4680
TATCCTGATT GGTGTTATTT ATGGTCTTAA AGAAGATAAG ATTATCTCTT CCTTCATGAA	4740
TGGTGCTGCT GACTTGCTCA GTGTGTCCTT GATCGTAGCG ATTGCTCGTG GTATTCAAGT	4800
TATCATGAAC GACGGTATGA TTACCGATAC AATCCTCAAC TGGGGTAAAG AAGGCTTGAG	4860
CGGTCTATCT TCACAAGTCT TTATCGTTGT AACTTATATC TTCTATCTAC CTATGTCATT	4920
CTTGATCCCA TCTTCATCTG GTCTTGCCAG CGCAACTATG GGTATCATGG CTCCACTTGG	4980
AGAATTTGTA AATGTCCGTC CTAGCTTGAT TATCACTGCT TACCAATCTG CTTCAGGTGT	5040
CTTGAACCTG ATTGCACCAA CATCTGGTAT TGTGATGGGA GCTCTTGACAC TTGGACGTAT	5100
CAACATTGGT ACTTGGTGGA AATTCATGGG CAAACTCGTA GTCGCTATTA TTGTAGTGAC	5160
CATCGCCCTT CTTCTCCTTG GAACCTTCCT TCCATTCCTA TAAATAGTG AGTGAGGTGA	5220
TTCCATGAAA ATAGATATAA CAAATCAAGT TAAAGATGAA TTTCTTATAT CATTAAAAAC	5280
CTTGATTGCC TATCCTTCAG TACTCAATGA AGGAGAAAAA GGAACACCTT TTGGACAAGC	5340
AATCCAAGAT GTCCTAGAAA AAACCTTAGA GATTTGTCCA GACATAGGTT TCACTACCTA	5400
TCTTGACCCT AAAGGTTATT ACGGATATGC AGAAATCGGT CAGGGAGCAG AGCTTCTGGC	5460
CATTCTCTGT CATTTGGATG TTGTTCCATC AGGTGATGAA GCAGATTGGC AGACACCGCC	5520
ATTGAAGCA ACTATCAAAG ACGGCTGGGT ATTCCGACGT GGTGTCCAAG ATGATAAAGG	5580
CCCTTCGCTC GCAGCTCTCT ATGCAGTAAA AAGCTTGCTG GACCAAGGTA TTCAGTTCAA	5640
AAAGCGCGTA CGCTTTATCT TTGGTACCGA TGAGGAAACC CTCTGGCGCT GCATGGCAGC	5700
CTACAATACC ATCGAAGAAC AGGCCAGTAT GGGCTTTGCA CCTGACTCAT CTTTTCTCTT	5760
GACCTATGCT GAAAAAGGGC TTCTACAGGT CAAACTTCAT GGCCTTGGAT CGGATCAACT	5820
AGAGCTTGAA GTAGGAGGCG CCTTTAACGT TGTACCAGAC AAGGCCAACT ACCAAGGTCT	5880
CCTCTATGAA CAGGTTTGTA ACGGTCTCAA AGAAGCTGGT TATGATTACC AAACCACTGA	5940

560

ACAAACCGTA	ACGGTTCTCG	GAGTGCCAAA	GCATGCTAAG	GATGCTAGTC	AAGGTATCAA	6000
TGCTGTGATC	CGACTAGCTA	CCATTCTTGC	TCCTCTCCAA	GAACACCCCTG	CTCTCAGTTT	6060
TCTTGCAACA	CAAGCAGGTC	AAGACGGCAC	AGGAAGACAA	ATCTTTGGTG	ATATAGCAGA	6120
TGAACCTTCT	GGTCACCTAT	CCTTTAATGT	CGCAGGTCTC	ATGATCAATC	ATGAACGTTT	6180
TGAAATCCGT	ATTGACATTG	GGACTCCTGT	CTTAGCTGAC	AAGGAAGAAC	TAGTAGAGTT	6240
GCTTACAAGA	TGTGCACAAA	ACTACCAACT	CCGCTACGAA	GAGTTTGACT	ATCTAGCGCC	6300
TCTATACGTC	GCAGAAGACA	GTAAACTCGT	TAGCACACTG	ATGCAAATCT	ACCAAGAAAA	6360
GACTGGCGAT	AACAGTCCTG	CTATTTATC	CGGTGGTGCC	ACTTTTGCTC	GCACCATGCC	6420
AAATTGTGTA	GCCTTCGGCG	CCTTATTTCC	AGGAGCGAAG	CAGACAGAAC	ATCAGGCAAA	6480
TGAATGTGCC	GTTCTAGAAG	ATTGTACCG	TGCTATGGAT	ATTTATGCCG	AAGCCGTCTA	6540
TCGACTTGCA	ACTTAATCAG	GCAACTGTTT	CTACCAAAAA	AAATCGACCG	ATTAATGAAC	6600
TGCACCCCAA	AAGTTAGACA	GAATAAATCT	AACTTTTGGG	GTGTTTTATT	ATGAAATTGA	6660
GTTATGAAGA	TAAAGTTCAG	ATCTATGAAC	TAAGAAAGCA	AGGACAAAGC	TTCAAACAGC	6720
TTTCAAAAAG	ATTTGGTGTG	GATGTTTCTG	GTCTAAAGTC	ATCTGAATCT	TTGAGATGAG	6780
CTTTATAAAT	CGCTTTTTTC	AGTTTTTGCA	CTGGTGTTTC	GATAAACTCA	AACTTTTTAG	6840
CCGTGGTATT	GCCTGATTTT	ATAGTATATT	GAAACTAGAA	TAGTACACCT	CTCCTTCTAA	6900
AACATTTTTA	GAAATCGATT	TGACTGTCCT	GATCGATTTG	TCCTGTTCTT	ATTTCATTTT	6960
ACTATATTTG	AGCCACTTCG	TCTTTAACGG	CTTTATTCAT	AAGCTCTTGT	AATTTTTCTT	7020
TACTATCAAT	TACTTCTGAT	TTCCGTTGT	AATTTATTGT	AATAGGTTTT	AACTTACCTA	7080
ATTTCTCGAC	ACGCTCATT	ATTTGATCTT	TTTTGAAGGC	TGCTTATGTT	TTTCCTAAGA	7140
TTTTTTCAAA	AATATAATTA	TCAGATAGCG	GTTTGCTCTC	TTCTTCAGCT	TGGTTTTTGT	7200
ATTAATTTGA	AACATAAGGA	ACAAATCCTT	CATAGTAACC	TAATGCTCCC	ATAAGTTCAA	7260
AAGCTTGTTT	TCTAATTCAA	ACCATTGCAA	CTCAGATTTT	AGCTTTTCAG	ATAAATCCTG	7320
CTCATCCAAA	TAATGACTTG	AAATTAGTGC	TGAACCTGTT	TCTGTATCCT	GTACAGGCTG	7380
AGCACCCATA	CCAGCAAAAA	ATAAACTCGT	TCCTAGCAAG	ACCGAACAAG	CTCCTATTGC	7440
ATATGGCCTC	AAAGAAAAAC	GCTGCTTTCT	CTCAAATTGA	AATTCTTTCA	TCCCATCTCC	7500
CATCATTCAT	TATTACTGTA	TATTTTGAT	ATCAGAAATA	GTTTGATTC	ACAAATCTTT	7560
CTAGTTATTC	CCTTATCATT	CCTAATTAAG	GGAGATAACA	TACAATAATT	TTTAGTTAAA	7620
TGTATATCGA	TGTTTTTTGT	TTTTCTTAAT	AAACGCAATA	CAAAAAGAGC	CTGTTACCAA	7680
GCTCTTTGTA	CTCAATGAAA	ATCAAAGAGC	AAATTAGGAA	ACTAGCCACA	GGTTGCTCAA	7740

561

AACACCGTTT TGAGGTTGCA GATAGAACTG ACGAAgTCAG CTCAAAACAC TGT TTTGAGG	7800
TTGCAGATAG AACTGACGAA GTCAGTAACA TCTATACGGC AAGGCGACGC TGACGTGGTT	7860
TGAAGAGATT TTCGAAGAGT ATTAGTCTAT TATTTCTTCT CAGCGCGAAG GGCTGACAAG	7920
ATTTGTGTTT GGATATCATC CACACCATTT GGAGTATTTG GTAAAAAGAT AGTTTGATTT	7980
CCTTTAGAGG CAAAGGTATT CAAGGTATCC AAATACTGGT TGGTCAAGAG GATAGACATG	8040
ATTTGTCTTT CTGTCATGCC AACATTGGCT TCCTTGAGTT CGGTGATAGA CTCTGCCAAT	8100
CCATCCACAA TCGCCTTACG TTGTTGGGCA ATCCCCACAC CATGAAGGCG GTCTTTTTCT	8160
GCTTCTGCTT CAGCTGCAGT GACAATTTTA ATCTTGTCAG CTTCGGCCAA TTCTTGTCGT	8220
CGGACCCGCT TACGTTGCGC CGCATTGATT TCATTTCATGG ATTGCTTAAC TTCTGCATCT	8280
GGTTCGACCT TGGTAATCAA GGTTTTCACG ATAATGTAGC CGTAAGTGGT CATTTCTTCT	8340
GCTACTTGGT GTTGAACCTC AAGGGCAATC TCATCTTTTT TCTCAAACAA TTCATCCAAG	8400
GTTAATTTTG GAACAGAAGA GCGAAGAGCA TCTTCGATAT AAGATTTAAT CTGAGATTCT	8460
GGACGTATGA GTTTATAGTA AGCATCTGTC ACGCTCTGCT CGTTGACACG GTAGTGAGTC	8520
GCTACATTCA TCATAACGAA CACATTGTCC TTGGTCTTAG TCTCAACCAC AATATCACTT	8580
TGCAACAAGC GCAACTGAAT CCGTGCTGCA ATCGAGTCAA TCCCAAAAGG CAAGCGAATA	8640
TGAATACCGC TATTAGCAAC CTTTGGTAT TTCCCAAAGC GTTCAATAAT CGCCACCGAC	8700
TGCTGACGAA CCACATAAAC TGTACTCAGT GTGACTATCA CCAATAGGAG CACACAAACA	8760
ATCAGAAAAA TCATGAAAAA TATTGCCATA ATGGAACCTC CACAAGTATT TTTCTAGTAT	8820
TATAGCACAT TTAAAGAAGG CTGTGCCGTT TTTACTGCGA TTTTCTCTGA AATGTCAATA	8880
ATTAGAGCTG AATTGTCTTA TTGTCGTCCA ATCTCTTGCT AAAATAACTC TTTATAAAAC	8940
GCAATCGTTT CTTCTAAGGT TGGCATAAAT GGATTTCTCT GTGCCGAGGC ATCAATCAAG	9000
GCATCTTAG AAAGGTATTC AAAGTCGAAA TCTTTTCTT CAATACCAAG TTCAGTCAGT	9060
TTCTTAGGAA TACCTACTGT CTCAGAAAGC TTCTCAATCT CAGCAATCGC ATAATCGGCA	9120
CATTCTTGAT CTGATTTACC TTCTACATGA AGTCCCAAGG CTTTGGCAAC ATTGCGGAAA	9180
GCTTCTGGTA CACGTTTAGC ATTTTCACGT TCTATAACTG GTAGCAACAT GGCACAGCAC	9240
ACGG	9244

(2) INFORMATION FOR SEQ ID NO: 69:

- (1) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 8898 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

562

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GATCTGAACT TTATCATCAT AACTTAATTT CATAATAAAA ACACCCCAAA AGTTAGATTT	60
TTTCTGTCTA ACTTTTGGGG TGTAGTTCAG TCATTGGACT GACGTTTTTT TGTATGCTTA	120
TTTGTATTTG ATGTAGTTGA TACCATCTGC TTTTGGTGCG ACTGCTTTTC CAAAGAAGGC	180
TGCTAAGACA AGAATTGTCA AAACATAAGG TGCAATTTGA AGATAAACCG CTGGCACTCC	240
TTGTAGGAAC GGCAATTGAG AACCGATAAC AGCCAAACTT TGTGAAAGTC CAAAGAAGAG	300
ACTAGAAAGC ATAGCACCGA TTGGATTCCA TTCCCAAAG ATCATCGCAG CAAGGGCGAT	360
AAATCCAGGT CCAACAATAG TTGTCACTGA GAAGTTAACT GAGATTGATT GCGCATAAAT	420
CGCTCCGCCA ATTCCACCTA GAAAACCTGA AATAATAACC CCTAAATATC TCATCTTGTA	480
GACGTGATT CCCAAGGTAT CCGCTGCTTG AGGATGTTCA CCCACAGAGC GGAGACGAAG	540
ACCAAATTGA GTCTTAAAGA GAATAAACCA AGCAAGGAAT GAGAAGGCAA TCGCCAGATA	600
ACCAAGTAGA CTAGTTGACT TGAAGAAGAT ATCACCATC ACTGGGATAT TTGCCAAGAC	660
TGGGAAATCA AAGCGTCCAA AAGTTTGACT TAGGTTGTCC GTTTGTCTT TGTTATAAAG	720
AACTTTAACT AAGAAAACAG CCAAGGCAGG CGCCATCAAG TTCAATACCG TACCGCTGAC	780
AACATGGTCT GCACGGAAAT GAACCGTCCG TGCTGCCGTG ATGATAGAGA AAACACTACC	840
AACCAATCCT GCTACAAGCA AGGATAGCCA TGGAGTTGCT GCTCCAAATT GTTCTGCAA	900
TTCAAGGTTA AAGACAACTC CAGAAAAGGC ACCCATAACC ATAATTCCTT CAAGGCCAAC	960
GTTTACCACA CCACCACGTT CAGAGAAAAC ACCACCGATA CTTGTAAAGA TGAGAGGTGC	1020
TGASTAAATC AGCATAGAAG ACACCAAGAG GGGGAGCAAG GTTATAATAG ACATCTTTAC	1080
TTACCTCCTT TAACTTGTTT TTTCGGTTTG ACAAAGCGTT CGATAAGGTA ATGAACACTG	1140
ACAAAGAAGA TAATAGACGC TGTTACAATG CTGACAAGCT CAGATGGTAC CTGCGCCGCA	1200
TTCATACCAG GAGCCCCAAC TTGGAGAACG CCAAATAGGA AGGCTGCAAA GAGTATACCA	1260
ATTGGTGAGT TGGCCGCAAG CAAACTAACC GCCATTCCGT TAAATCCGAT AGCTAATGAC	1320
GAACCTTGAA CATAGACGTT CTGGAAGGTT CCCAAACCTT CAACAGCTCC ACCAAGACCT	1380
GCCAAGGCAC CTGAAATAAT CATAGATAGG ATAATAGTCC GCTTGGCAGA AATACCAGCA	1440
TATTCTGAAG CATGTGGATT AAGACCAACT GCACGGATT CAAAACCAAG AGTTGTTTTTC	1500
TTGAGCATGA ACCAAATAAC TGCAACGGCA ATGATGGCAA AGAAAATACC AATATTCATC	1560
CGTGAGTTAC CAGTCAACTC AGCCAACCAA GGTGTCTGAT AGGTTGCATT AGCCCCAACA	1620



563

CGAATGGTCG AATCTGTACT TTGCATGAAG TCTTTAGGGA AAGCATGGAT AAAGGCATTG	1680
CCTACATACA AGACAATGTA GTTCATCATG ATGGTTACAA TAACCTCTGA CGTCCCTAGA	1740
TAGGCCCTAA GAATACCTGG AATCGCTCCG ACAATCCCAC CAGCAATCAA GGCAATCACC	1800
ATGGTTGCTA GAATCATCAA GGGACGGGGC ATATCTGGAT GCGACAGGGC AAACCAACCA	1860
CTGAGAATCC AACCTGCCAA AGCCTGACCA GGAAGTCCGA CGTTAAAGAA ACCAGCTCGA	1920
CTGGCAACGG CAAAACCAAG ACCAATCAAG ACCAGAGGAC CCATAGCAGG GAAGATTTCT	1980
CCAATCCCAC GCAGACTGCC AAAGGCTGTA TAGAACAATT CTTCGTAGCC CCAAATAGCA	2040
TCATAACCGA AGATCCACAT GACAATGGCT CCGAGTAAAA TTCCTAGGAA TACAGAAATC	2100
AAGGGAACCG AAATTTGTTG TAATTTTTTA GACATCACTC TTCTCCTTTC CCAAGTTTCC	2160
ACCAGCCATC AAGACACCAA GTTCTTGTTT ATTGGTTGTT TCTGGTGATA CAATACCTTG	2220
AATCTTACCA TCGTGGATAA CGGCAATACG GTCTGAGACG TTTAAATCT CATCCAATTC	2280
AAAGCTGACA ACAAGGACAG CCTTGCCATT ATCACGCTCT TCAATCAAGC GTTTGTGGAT	2340
ATACTCAATG GCACCGACAT CCAACCCACG AGTTGGCTGG CTAACGATAA GGAGATCAGG	2400
ATCTCGATCA ATTTACAGAG CAATAATTGC TTTTGTGTA TTTCTCCTG AGAGTGCAGC	2460
TGCAGGAAC TTTCACTGG CAGCGGAAC ATCAAATCT TCCATCAGCT TTTTAGCATA	2520
AGAAGTAATA TTTGAATAAT TCAAAATTC ATTTTACTA TGTGGTTCTT TATAGTAGGT	2580
TTGAAGGGCA ATATTTTCAG ATATCATCAT TTCCAAAATC AAGCCATCAC GGTGACGGTC	2640
TTCTGGAACG TGCCCAACAC TTAGTTCTGT AATCTGACGT GGGTGCAAGC CTACAATTGA	2700
ATCTCCTTTT AGCTCAATGC TACCAGATTC AACCTTACGA AGACCTGTAA TGGCTTGAAT	2760
CAGTTCAGAC TGACCATTTT CATCAATCCC CGCAATACCA ACAATCTCTC CAGCACGAAC	2820
ATCCAAGGAC AGATTTTAA CAGCTGGAAC ACCACGGTTT TCATTGACCA CCAATCTTT	2880
GATAGACAAA ACCACTTCTT TTGGTTTAGA GGCTTGCTTC TCTGTTTTAA AGGAAACAGA	2940
ACGTCCTACC ATCATTTCCG CCAAATCAGC ATTGGTAGCC CCTGCAATTT CAACGGTTTC	3000
AATTGATTTT CCACGACGGA TAACTGTAAC ACGGTCAGAA ACTGCTCGAA TTTATCCAA	3060
TTTGTGGGTA ATCAAGATAA TTGATTTTCC TTCTTTGACA AGATTTTCA TAATAGCCAT	3120
CAACTCATCA ATTTCTGATG GAGTCAAAAC AGCCGTTGGT TCGTCAAAGA TAAGGATATC	3180
AGCCCCCGA TAAAGTGTTC TTAATTTTC TACAGTTGT TGGGCTCCAA CTGAGATATC	3240
TGCTACCTTG GCAGAAGGGT CAACAGCTAA GCCATAACGT TCAGAAAGAG CCTTGATTTT	3300
TTTGCTAGCT CCAGCGATAT CTAGCACACC ATTTTATGTC AATTCATAC TAAAATGAT	3360

564

GTTTTTCAGCC ACTGTGAAGG CTTCAACCAA CATAAAGTGC TGGTGAACCA TCCCGATTCC	3420
CAAGCTAGCT GCTTTAGATG GGGAGTCGAG ATTGACAACCT TGACCGTTGA CCGCGATTTC	3480
ACCACTAGTT GGTTCAGAA GGCCTGCTAA CATGTTTCATT AGCGTGGACT TACCAGCCCC	3540
ATTTTCTCCT AAAAGTGCAT GAATTTTACC TTTTCGTAGG TGCAAGTTGA TTTTGTCTGT	3600
GGCAACAAAT CCACCAACA CCTTGGTAAT ATCAGGCATC TCAATGACAT TTTCTGTGTC	3660
CATGTGCTCT TCCTTTCAGA GTCTTATTTT ATTTCAATAA AACTTGCTAG TTTGTCTAGT	3720
AGCAAGCTTT ACTTAGACAA AATGACTTTG TCTCAACTCT TAAAAAAGCG GCCCTTGGCC	3780
GCTTCCTAAG AAATGACTTC CATCCATTAT TTTTCAGGAA CTPTTACGCT TCCATCAAGG	3840
ATTTTAGCTT TTGCATCTC GACAGCTTTT TTACCTTCTT CTGAAAGGTT TGTACTGCC	3900
AAGTCAACCC CTTTATCCTT CAATGAGTAA ACGATCACTT GACCGCCAGG GAATTCTCCT	3960
CTTTCTGCCT TGTTAGAAAT ATCTTTTACA GTTGTACCAA CTGTTTCAA AGTAGATACA	4020
AGAACAAAGT TTGATTCTTT GCCATCTTTA GAAGTGTATT TACCTTCTGC TTCTTGCTCA	4080
CGATCAACAC CGATAACCCA AACTTTTTCA TTTTCAGGAC GGCTTTCGTT GAGAGATTTT	4140
GCCTCTGCAA AGACACCTGC ACCTGTACCA CCAGTACTT GTTAAACAAT ATCTGCACCG	4200
GCTGCGTATT GTGCGGCTGC AATTGTTTTA CCTTTAGCCG CATCACCAA TGAACCAGCG	4260
TAGTCAACTT GGACTTTGAT AGATGGGTCT ACTGACGCAA CACCAGCCTT GAATCCTGCT	4320
TCAAAACGAG AGATAACTTC AGATTCGATA CCACCTACAA AACCAACTTG TTTTGTCTTA	4380
GTTGTTTTTG CTGCAGCCAC ACCTGCAAGG TAACCTGACT CATTATCAGC GAAAGTTACG	4440
CTCGCAACAT TCTTTTGCTC TTTAATCACA TCATCAATCA AGACATAGT CAAGTCAGTG	4500
TGTTCTTTTG CTGCATCTTT AACTGCATTA TTAAGGGCAA AACCAACACC GAAGATTAGG	4560
TTGTAACCTC CAGCCGCTTG TTGCAAGTTG TTAGCGTAGT CAGCTTCAGT TCTTGATTGG	4620
AAGTAAGTGA AACCGTTATC TTTTGAAAGA TTGTGTTCTT TACCCCAAGC CTGCAAACCT	4680
TCCCAAGCTG ATTGGTTGAA TGATTTGTCA TCAACACCAC CAGTATCAGT GACGATTGCT	4740
GCTTTTGTCT TCACATCAGA AGATGAAGCT GCGTTACGAG AAGAGCGGTT ACCACATGCA	4800
GCAAGTCCAA CTGCTGCCAC TGCAACTAGG CCAAGACCTA GCCATTGTTT CTTGTTTCATT	4860
ACTGAACCTC CTAAATAAGA TGTGCAACGA TGTTGCAAGT ATGGATTGGT TGGCCACAAG	4920
GACCGTGCCA CTCAGAGAGC GACTCAGACT AGTTTAAGTC TGTAAGAGAG TATGGAAGTA	4980
ATTCCCCGAC CGTCATCTCG ACCGTCGATT TATCTTTTGC GACTAAGGTC ACTTTTAGAT	5040
CTTGTTCAAA AAATTCAGCC ATCACTTGGC GACAAGCACC ACATGGCGAG ATCGGTTTTT	5100
CAGTTTGACC ATAGACAATC AATTCTGAAA ATTCTCTTTG GCCTTCAGAT ATAGCCTTAA	5160

565

AAATAGCTGT TCTCTCACCG CAATTGGTCA AAGGATAGCT AGCATTTTCA ATATTCACCTC	5220
CCGTGTAAAC ACTTCCGTCT TTAGCTACTA AAAGTCTCC GATAGGAAAG TGAGAATAGG	5280
GGACATAGGC ATGTTTGCTG GTTTCAATTG CCAGTTCAAT CAACTCAGTA GTCGCCATCT	5340
GCCAATTCTC CTTTAAAAAT AGCTACCCCA GCTGACGTTG CGATACGGGT CGCACCTGCT	5400
TGACAAAGG CAAGAGCATC TGCATAAGAA CGAGCTCCAC CGGCGGCCTT GACACCCATA	5460
TCAGATCCAA CTGTTTCACG CATTAAATGA ACATCTGCTA TCGTAGCACC ACCAGTTGAA	5520
AAGCCAGTAG ATGTTTTGAC AAAGTCAGCC CCAGCTTTTT GGGCCAATTG GCAAACAACA	5580
ACTTTTCTTT GGTCTGTCAG AAGGCAAGCT TCAATAATGA CTTTCACTAA CTTATCACCA	5640
CTTGCTTCCA CTACTGCGG AATATCTGAC TCAACCAAGG CTAAATTACC TGATTGAGA	5700
GCTCCAACAT TGATCACCAT ATCAATCTCA TCTGCACCAT TTTGGATAGC TTCTTTTGTC	5760
TCAAATGCTT TCACGGCTGA AGTTGTTGCT CCCAAAGGGA AACCTACTAC TGTGCAAACC	5820
TTAACATCTG TGCCTTCAAG TCCTTTTTTA GCATGTTCAA CCCAGGTCGG ATTAACGCAA	5880
ACACTGGCAA AGTCATACTC TCTAGCCTCA GACAACAAAC TATCAATTG TTTTTCTTT	5940
GCATCTTGTT TTAAGAGCGT ATGATCTATA TATTTATTTA ATTTCAATTG GGTTCCTCT	6000
CCATTAGGA GATGATTTCT ACAATTCAC GGATTTTTTT CACTTCATCA CTTATTTTAA	6060
CACATTTTGG GAAATCTGTA ACTAGTTGAG GTGGAATTTT TTCATTTGTG TATACTTTTG	6120
CAACAATTTT ACCCTTTTGA ACGGAGTCTC CAATCTTCTT TTCAAAAACA ATTCCTGTTT	6180
CATAGTCCAA GGCATCAGAC TTAAGTGCAC GACCAGCACC CAGCCTCATG GCATAAGAC	6240
CAAAGTCCAT AGCTGGAAGA GCTGAAATGA CACCCGTTTC CTGAGCAGGG ATTTCCACCA	6300
CATGAGCTAC ATTTACAGGA CGATAGAGGT CTTCCAAGTC TCCACCTTGG GCTTGACCA	6360
TTTCCTCAAA CTTAGCCAGT GCTTGACCAT TCTCAAGATG TTGGTGAAGT TCTTCAACAG	6420
TTTTGTAAAC ATTTGCCAAA CCAAGCATAA TTTGAGCCAA TTCACAAATA AAGTGGGTAA	6480
TATCTGACG TCCTTGACCT TGCAAAATCT CCAATGCTTC AAGGATTTCC AGACGATTTT	6540
CAATCGCTCG TCCCAAAGGC TGGTCATAT CCGTAATCAC TGCTACTGTC TTCCGTCCAA	6600
CAACCTTACC AAGATCTACC ATAGTTTGAG CCAACTCACG CGCCTCATCA ACCGCTTCA	6660
TGAAGGCACC CTCACCGACA GTCACGTCTA GCAAAATAGC ATCCGCCCCCT GCCGCAATTT	6720
TCTTGCTCAT CACCGAACTC GCAATCAAAG GAATCGTGTC GACAGTTGCG GTCACATCAC	6780
GAAGGGCATA GAGAAGCTTA TCTGCTTTGA CCAGCTGGTC TGATTGCCCA ATGACAGATA	6840
CTCCAATATC CTGAACCTGA CGAATAAAAT CCTCTTGACT ACGTTCTACT TGATAGCCCT	6900

566

TAATGGACTC	CAATTTATCA	ATTGTTCCGC	CTGTATGGCC	AAGACCACGA	CCACTCATTT	6960
TTGCTACAGG	CACACCGAAG	CTAGCAACAA	GAGGAGCTAA	AATCAAGGTT	ACCTTATCGC	7020
CGACACCACC	AGTAGAATGC	TTGTCAACTT	TCACACCATC	AATGGCTGAC	AGGTCAAACCT	7080
CTTGCCCGAGT	CTTAACCATA	TTCATCGTTA	AATCAGAGAT	TTCTCGAGTC	GTCAATTCCTT	7140
TAAAATAAAC	AGCCATAGCA	AAGGCAGACA	TCTGATAATC	AGGAACAGTT	CCTGATACAT	7200
AGCCTTCTAT	CAGCCATTCA	ATTTCACTTG	AAGTCAGTTC	TTGACCGTCT	CGTTTTTTTT	7260
GGATTAAATC	AACTGCTCTC	ATTCTTTCAC	ACTTCTAAGG	ATATAGTATC	CCTTGCTCTT	7320
TTTAAGGATT	TCACAATTGC	CAACACATC	TTCCATCTTA	GACTTGGCAC	TTGGAGCTCC	7380
TTGTTTTTTC	TGGATGACGA	TGGTCAAATC	TCCACCAATT	TCCAAGAAAT	CTTTACTTTT	7440
CTCGATGATT	TCATGAACGA	CTTGCTTGCC	CGCACGGATA	GGAGGATTGG	AAATGACATG	7500
GTCAAATCGC	CCTTGAACCT	TTGCATAAAT	ATTAGATTGA	AATATCGTCG	CTTTTGCATT	7560
ATTTTTTTCA	GCATTTCTCT	GAGCTAAATC	CAGGGCACGA	GTGTTAATAT	CAACCATGGT	7620
CGCCTGAACT	CCGTAAACCT	TGACCAAGGA	CAACCTAAT	GGACCATAAC	CACAGCCTAC	7680
ATCTAGGACT	GTCTCTCCTT	GGTTGACATC	CAGACACTTG	AGCAAGAGTT	GACTTCCAAA	7740
GTCAACCATT	TTCTTGCTAA	AAACACCCGC	ATCTGTCAAA	AAAGTCATTT	TTTCTCCCAA	7800
CAAGTCCACT	CTCAACTCAT	GAATGTCGTG	AGCAGCGTCA	GGATTTTCTG	CATAGTACAT	7860
TTTACTCATG	ACACTATTTT	ACCATAATTT	GACTCAAATT	GTAAATCGTT	TACAAATTGA	7920
TAATAAAACG	AAAAAGACCG	AAGAAAGCAA	GTACGAAGC	CATTTTCTTC	AATCTCTTTC	7980
AACACTTATA	AATAATAAAC	CATTTAGAAC	TATAAATATC	ACAGTCCAGA	TAAAAACAAA	8040
AAGTTTATCA	TCTATAATCA	GGCAGATTAT	TATTTCTATT	GCTTAACCTT	AAAATACTTT	8100
ATTATCAACA	AAATTCCTAA	CAAAATGTTT	AGATAAAAAGC	CCAACTGATA	CGTTTATGTC	8160
AGGATTTCCA	AACTTGTTCA	AAGTCGTATC	AAATCTTCTA	GTGACATGTG	GAAGAAATAA	8220
CCCTCTGTCTG	CAATCCGTAG	GACTAAAAAG	CAATAACTAC	CCGCAGCAAT	CCATTTCGTCT	8280
CATCGTTTTT	TAGTAAGAAA	GCAATTAAGA	ACGAACAAAT	AAAGACAGCT	GTTACAATAG	8340
CATGTTCCAT	CAAAAAAGTA	AAACCGTAAT	AGGTTTCCAC	AAAGCATCTA	CCATTATCTG	8400
CATTGGTTCC	TTTTATAAAA	GGTAAAGCAA	AACTTAAAT	AAAACAGAGT	TCCAATATGT	8460
AACGTTTTAA	GATTTTCATA	GTACACCTCC	TATAAGTTGT	GAACTAAAAA	GCCCCCTTTA	8520
TAAGCTTATA	AATCAGTAGA	ATCTATCTCC	TATTTATCA	ATAAATTGAT	CACCTTATACT	8580
ATATACCATT	GACTTACCAC	ATTCAAGAAA	CCGCTTTATT	TTTTTAGCTT	TTTATGGTAT	8640
GATAGACAAA	ATATCTAGGG	GAAAAACAAAT	GACCAACGAA	TTTTTACATT	TTGAAAAAAT	8700

567

CAGCCGCCAG ACTTGGCAAT CTTTACATCG AAAGACAACA CCTCCTTTGA CAGAAGAAGA	8760
ATTGGAATCT ATCAAGAGTT TTAATGACCA AATCAGTCTC CAAGACGTTA CAGATATCTA	8820
TCTCCCCTTG GCTCATTGTA TTCAGATTTA CAAGCGAACT AAGGAAGATT TAGCCTTTTC	8880
AAAAGGAATT TTCCTCCA	8898

## (2) INFORMATION FOR SEQ ID NO: 70:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13188 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TATCTTAACG aGGATTGGGT TTATCGTCAG TCTTATTGCC CTAATTGTGG GAACAATCCC	60
TTAAATCATT TTGAAAATAA TCGGCCTGTA GCAGATTTTT ACTGTAATCA TTGTAGTGAG	120
GAGTTTGAAC TAAAGAGCAA AAAAGGAAAT TTTTCATCAA CAATCAATGA TGGTGCCTAT	180
GCAACGATGA TGAAGCGTGT GCAGGCAGAT AATAATCCTA ATTTCTTTTT TTAACTTAC	240
ACAAAAAATT TTGAGGTAAA TAACTTTCCTT GTCCTTCCGA AGCAATTTGT TACACCGAAA	300
TCGATTATTC AAAGAAAACC ACTTGCACCA ACTGCTAGAC GAGCAGGTTG GATTGGTTGT	360
AACATTGATT TATCACAAGT ACCTTCTAAA GGAAGGATAT TTCTTGTGCA AGATGGACAA	420
GTTAGAGATC CAGAAAAAGT TACAAAAGAA TTTAAGCAAG GTTTATTTTT AAGGAAGAGC	480
TCTCTGTCAT CAAGAGGTTG GACAATAGAA ATTCTAAATT GTATAGATAA GATAGAGGTT	540
TCAGAATTTA CCCTTGAAGA TATGTATCGT TTTGAAAGTG ACCTAAAAAA TATCTTTGTT	600
AAGAACAATC ATATCAAAGA AAAGATTAGG CAACAGCTTC AAATATTAAG AGACAAAGAA	660
ATAATAGAAT TTAAAGGTAG AGGAAAGTAT CGGAAATTAT GAAAACGAAA CAACTTGTG	720
CATCAGAAGA GGTGTATGAT TTCTTAAAAG TCATCTGCCC TGATTATGAA ACTGAAAGCC	780
GTTACGATAA CCTAAGTTTA ATCGTCTGTA CTTATCAGA TCCCGATTGT GTGAGATCGT	840
TATCTGAAAA TATGAAATTT GGTGACGAAA AACAACTAGC TTTGATGAAG GAAAAATATG	900
GGTGGGAAGT AGGAGATAAA TTGCCAGAGT GGCTACATAG CTCCTATCAT AGATTATTGT	960
TAATAGGTGA ATTATTGGAA AGCAATCTAA AACTGAAAAA GTATACAGTA GAAATTACAG	1020
AACTTTATC ACGTTTAGTA AGTATAGAGG CTGAAAAATCC AGATGAAGCC GAACGACTTG	1080
TAAGAGAAAA GTATAAGAGT TGTGAAATTG TTCTTGATGC AGATGATTTT CAGGACTATG	1140

568

ACACTAGCAT ATATGAATAG GTAGATGTTT TTATTTTGTC AACAAAAAAG AGGCTCGCAC	1200
CTCTTTTCT TATTTCTTTT TATGATTAA TACGGCATTG AGGACAATAG CGAGTAGGCT	1260
GGCTACGACG ATTCGGTTTG AGAAGAACAT TTGGAAGGCT GTCGGCATGC TGACAAAGAG	1320
ATTACTGTTG TTGAGACCGA CACCTGCAGC GATTGAAACA GCTGCGATAA GGAAGTTGTG	1380
TTCATTGTTA GCAAAGTCAA CACGGGCGAG GATTGTCATC CTTGAATTG ATACAAAACC	1440
AAACATTACC AGCATGGCAC CACCGAGGAC GGAGCTTGGG ATGATTTGGG CAAGGGCGCC	1500
AAACTTAGGA AGCAGTCCAA GGAGAACCAG GAAACCAGCT GCGTAGTAGA TTGGCAGGCG	1560
TTTTTTGATG CCGTACAATT TAACCAAACC AACGTTTTGT GAAAATCCGG TGTAAGGGAA	1620
GGTGTTAAAG ATTCTCCGA GAAGTACGGC CAAACCTTCT GCGCGGTATC CGTTGCGAAG	1680
GCGCGTGCTG TCGATTGGAT CTTTGTGAT ATCAGACAAG GCCAGATAAA CACCAGTTGA	1740
CTCAACCATA GACACCGTTG CGATGATACA CATCATGACA ATAGATGAGA TTTCAAAGGT	1800
TGGCATCCCA AAGTAGAGTG GAGTTGGGAC ATGGACAAGT GGAGCTACCG CAACAGGAGA	1860
GAAGTCCACC AAGCCCATAG TAGCAGCAAT GGCAGTTCCA ACAACCAGAC CAATCAAAAT	1920
AGAGATAGAC TTGATAAATC CTTTGGTAAA GATGTTGATC AAGAGGATAA TCAGAACAGT	1980
AATAGCTGCA AGCAAGAGAC TTTGACCAGT TGGCTCTGGA ACGTTATTTT CCATATTTCC	2040
AATAGCGACA GGGATCAAGG TTAACCAAT CGTGGAATA ACAGATCCTG TTACGATAGA	2100
TGGGAAGAGA TTGGCTACTT TTGAGAAGAT GCCTGAAACA AGAACCACGT AAATCCCAGA	2160
TGCGATAAGG GCACCAAACA TAGCGCCACT ACCATGGCTT TGCCCAATCA TAATCAAGGG	2220
AGCGACCGAC TGGAATGCAA CTCGAAGAAC GACTGGGAGT CCAATCCCAA AGTATTTGTT	2280
GAGTTGGAGT TGGAGGAAGG TTGCCACCCC ACACATGAAG ATATCTGTAG AAATCAGGTA	2340
GGTCAACTGC TCAGCTGAAT AGCCAAGGSC TGTGCAATC ATGATGGGAA CCAGGATAGA	2400
TCCTGAGTAC ATGGCTAGTA AGTGCTGCAA GCCAAGAACG GCTGCTTGGC AGTGTTTTTC	2460
TTGAGTTTGC ATTAGAGATC TGCCTCCTTA AATACGACTT GACCATTTTC AAAACAATCC	2520
AAACGAGCAA GTGATAGGAC AGGGTAGCCT GCTTTTTCAA GCAAATCACG ACCATCTTGG	2580
AAGGATTTCT CAATCACGAT ACCGATAGCT TGGACTGTGG CACCGGCTG TTCGATGATT	2640
TGAATCAAGC CTTTAGCAGC TTGGCCATTA GCAAGGAAAT CGTCGATAAT CAAAACCTTG	2700
TCCTCTGGTG AGAGGAATTT TTCAGCGATA GAAACGGTGC TGGTCACCTG CTTGGTAAAG	2760
GAGTAGACTT GAGCAGTTAA GATGCCTTCG TTCATGGTGA TGTCTTAGC TTTTGGCG	2820
AAAATCATGG GAACGTTTAA GGCTTCAGCT GTAAAAACGG CTGGGGCAAT ACCCGACGCT	2880
TCAATGGTTA CGACCTTGGT AATGCCAGTA GTAGCAAAT TTTCCGCAAA AACCTTACCA	2940

569

ATCTCTCGCA TCAAGCTAAA GTCAACTTGG TGGGTAAAA AGGAATCTAC CTTGAGGATG	3000
TTATCACCCA AGATATGCCC ATCCTTGAGG ATGCGCTCTT CTAATAATTT CATAAGACCT	3060
CCTAAAGTCT AAAAGTTAAT TTACTTGTG TTTAAATATT TCTATAGTGA TCCCTTTTGC	3120
TAATACTATA TATTTGATAA AACTATTACG AGCGAAGCGA GTCTTATCAA ATATTTCCCG	3180
TTGTAGTGGT ATCATAGACA ATAATCTTGT TATTGTCTAT GACGGGATTT TTGAGAGTAA	3240
AATAGTTCCG GGAAGTATTT TAGCCTAAGC CTAGAAATGA AAGAGCTAGG GGCTCAAAAA	3300
TTAGGGATGA AATTCCTGG ATTCTGAAA TTATTCACAG GATAATTTCA CCTCCCGTCC	3360
GCACTAATTA AGGGAAATAT TAAAAAAGA CCTACTTAAT CTCTAAGTAA GTCCCTTAAA	3420
TAGACATGGC AAAAACGGCC ATATCTCACT GCTGACTTAC TTATTGTTAG GTGTTCCGGC	3480
ACCTTGTAGA AACGTCGTGC CAATTCACGA CATAACAAG TAAAACGATA TTCAATTTTA	3540
AATAGGCTTG AGCCAATGTT TTTATTTTAC ACTAAATAAC TTTAGAAATC AACTATTTTG	3600
TTAGTCTTTT GGTTTAAAAA ACGAACAAAA AGAAGAGAGG GTGAACAAAA ACTCCATTGT	3660
AAGCTAACAG TTATACTAAA TGAAAATCAA AGAGCAAACT AGGAAGCTAT CCACAACCTC	3720
AAAACACTGT TTTGAGGTTG TGGATAGAAT TGACAGAGCC AGTATCATAT ACCTACGGTA	3780
AGGCGACGTT GACGTGCCTT GAAGAGATTT TCGAAGAGTA TTAGAAGATT TTTCCATCAT	3840
AAAAGGCATA CTATCAAGCT TTTAGACACC TGACAATATG CCTTTTCTA ACTTTAAAGA	3900
CTTTTCCCAA TTTTATTAT TCTACTCGCT AAATCTTAAA AAATAGCCAT CTGGATCCAA	3960
AACTGCAAAT TTATGAGGAT AGATATAGGG ATCACTGACA CGAAACTTTC TTTTGGTCAA	4020
GGGACGATAA ATAGGATAGT TTGCCCTCAT CACTCTTTAA TAGAGTTTGA AAACATCCTT	4080
TATGCCAAAG GAGAGATTGA CTCCACGACC AAAGGGATAG GTCAGTTCAG CTAGTTGATC	4140
CTTTGTTCCC TCCTCTAACA TTAGTTGACA CTCTTCAAGA GAAAGAGAAA GTTTTCTTCT	4200
GGACGTTGGT ATTCAATCCT AAAACCCAGT AAACCACAGT AGAAGGACCG GGAAGTTCCG	4260
ATATTCGATA CAAGCAACTC GGAATGACC GCATTGTAGT CCATATAGAA AATCCTTACA	4320
AGTCAATTTT CAAGACAATC GGTGTATGGT CTGGCGAGC ACCTGAGTCA ATCATATCAG	4380
ATTTAGTGAC CTTGTCAGCG ATACGGTTAC TTGTGAGCCA GTAGTCGATT CTCCAGCCTG	4440
TATTGTTGAT TTTAGAAGTT TTGCTGCGTT GTGCCACCA AGTGTAGCGT TCAGGAACAT	4500
CGCCATGAAC ATGGCGGAAG GTGTCTGTAA ATCCAGTTGC CAAAAGGTTG GTAAATCCAG	4560
CACGTTCCCTC GTCAGTAAAT CCAGGTGAAC GCGGTTGCT AGCAGGATTT GCAAGGTCGA	4620
TTTCATTGTG GGCTACGTTG TAGTCACCGG TCGCAAGGAC TGGTTTTTCT TTGTCTAGTT	4680

570

CAGCCAAATA CTCAGCATAT TTGGCATCCC AGACTTGGCG TTCTTCCAAG CGTTTGAGAC	4740
CGTCACCAGC GTTTGGAGTG TAACTTGGG TTACGAAAAA TGCAATCAAT TCTAGAGTGA	4800
TGATACGACC TTCCAAGTCC ATGGTAGAAG GGGCACCGAT TTCTGGGAAG CTGATAGTAG	4860
GTGTAAGTTC TTTCTTATAA AGGAACATGG TTCCAGCATA GCCTTTACGG GCAGGCTCTT	4920
GGGAAGAGCG CCACGTGTTT TCGTAGCCTG GGAAGAGTTC TTCTAAAATT TCCACGTGTT	4980
TCTTTGTAGG TCCTTTGGCA GAAAGCTTGG TTTCTTGGAT AGCAATGATA TCAGCATTTT	5040
CAGCGACCAA GGTTTGTAGG ACTTCTTGGG ACAATTTGGC ACGAGCTGAG TCACTAGTTA	5100
GGGCAGCGTT TAGGGAATCA ATATTCCATG AGATAAGTTT CATAAAGTTA CCTTTTTTCAT	5160
TCAGATTATA GATTTTATTA TACCAAAAAA AGATCTATTT CCCCAACGTA TGGTTTGAAA	5220
AATTACTCTC TTTCTTTTAT AATTAAGAAT GATTTTATGA AAGGGAGTGA AAATACATGA	5280
AATTCTACTC TTATGACTAT GACTCAGCC AAATCGGTCA GCAAAATGGT ATCATGGTTG	5340
GCTTTGGGAT TGTTCTATTA CCTGTGACAG TTTTTTTTGC TTTCAAGGCA TACCATAATA	5400
AAAAGGGAAG CGAATTTCTG GAGTTGGTCA TGATTTCAGA TCTGGCCTTA TTTAGCTCTG	5460
CTTTTGGTCA GCATCAGGAC TTATCAAAAC AATCAAGTTT CTAACAATAA ATTTCAAAC	5520
TCACTTCATT TCATCGAGGT TGTTTCCAAA GATTTGTGAG TAGACAAGTC AGAAGCTAT	5580
GTTAATACTT CCACAAACAC AGATGGCGCA CTTATCAAGG TGGGAGATCG CTATTATCGT	5640
GCCCTAAATG GAAGTGAGCC AGACAAGTAC CTGTTAGAGA AAGTCCAATT GTATAAGACA	5700
GACGCAATTG AACTGGTGGG TGTGAACAAA TGACACTTAA TTATATCGAA ATTTTAATCA	5760
AACTGGTCTT GACTCTCAAA TAGCTCAACA ACAATGTTCA CTTTGTGAAA CGTTTGATTG	5820
ATGGTAAGCC AACTCTCCTT ATCAAAAATG GGAATATTGA CCCAGAAGCC TGTGCTCAG	5880
TTGGTTTGTC TGCATCGGAT GTATCCCTCA AACTTCGTAG CCAAGGGATT TTCCAGATGA	5940
AGCAAGTCAA ACGAGCTGTG CAAGAGCAAA ATGGGCAACT CATCGTTCTG CAAATGGGAG	6000
ATGAAAATCC TAAGTATCCA GTTGTGACTG ACGGTGTGAT TCAAGTAGAT GTCTTGAAT	6060
CGATTGGTCG TAGCGAAGAG TGGTTGCTTG ATAACCTCAG TAAACAAGGG CATGACAATG	6120
TAGCCAATAT CTTTATTGCT GAATATGACA AGGGTGCTGT TACAGTCGTA ACTTATGAAT	6180
AAGAAAAACC TGGGGTCTTG TACTCTTCGA AAATCTCTTC AAACCGCGTC AACGTCGCT	6240
TGCCGTATGT AGGTTACTGA CTTCTGTCAGT TCTATCTACA ACCTCAAAGC AGTGCTTTGA	6300
GCAGCCTGCG GCTAGTTTCC TAGTTTGCTC TTTGATTTTC ATTGAGTATT GGCCTCAGGT	6360
TTCCATTTGC AATCAGAAAG GGATTTTATG TCCATTATTC AAAAATTTG GTGGTTTTTC	6420
AAGTTAGAAA AACGCCGTTA TCTAGTCGGA ATTGTGGCCC TGATCTTGGT TTCCGTCCTC	6480



571

AATCTCATTCTCCTATGGT TATGGGGCGG GTCATTGATG CCATCACATC GGGGCAATTA	6540
ACCCAGCAGG ACCTCCTTCT TAGCCTATTT TACTTGCTAC TTGCAGCCTT TGGTATGTAC	6600
TATTTGCGCT ATGTGTGGCG TATGTATATC CTTGGGACCT CTTATTGCTT GGGACAGATC	6660
ATGCGGTCTC GCTTGTTTAA GCATTTTACA AAAATGTCGT CAGCCTTTTA TCAAACCTAT	6720
CGGACGGGTG ATCTGATGGC ACACGCAACC AATGATATCA ATGCCTTGAC TCGTTTAGCA	6780
GGTGGCGGTG TCATGTCTGC GGTGGATGCC TCTATCACGG CTCTGGTGAC TTTGTTGACC	6840
ATGCTCTTTA GCATCTCATG GCAGATGACT CTTGTTGCCA TTCTCCCCCT ACCTTTCATG	6900
GCCTATACGA CTACTCGCCT AGGGAGAAAAG ACTCATAAGG CCTTTGGCGA ATCCCAAGCT	6960
GCTTTTCTG AACTCAATAA CAAGGTACAG GAGTCCGTAT CAGGTATCAA AGTGACCAAG	7020
TCTTTCGGTT ATCAGGCAGA CGAGTTGAAG TCTTTTCAGG CAGTCAATGA ATTAACCTTC	7080
CAAAAGAACC TGCAAACCAT GAAATATGAT AGTCTCTTTG ACCCTATGGT TCTCTTGTTT	7140
GTGGTTTCGT CCTATGTTTT AACGCTTTTG GTTGGCTCCT TGATGGTTCA GGAAGGGCAG	7200
ATTACAGTTG GGAATCTAGT CACCTTTATC AGCTATTGG ATATGCTGGT CTGGCCTCTT	7260
CTGGCCATCG GTTCTCTCTT TAATACTACT CAGCGAGGGA AGGTTTCTTA CCAGCGGATT	7320
GAAAATCTTT TGTCTCAGGA ATCTCCTGTA CAAGACCCTG AGTTTCTCTT GGATGGTATT	7380
GAAAATGGGC GTTTGGAGTA TGCCATTGAC AGCTTTGCTT TTGAAAATGA GGAAACACTG	7440
ACGGATATTC ACTTTAGTTT GGCAAAAGGG CAAACACTGG GCTTGGTTGG GCAGACAGGC	7500
TCTGGGAAAA CGTCCTTAAT CAAGCTCCTC TTGCGTGAAT ACGATGTGGA TAAGGGTGCC	7560
ATTTATCTAA ACGGTCACGA TATCGGGAC TATCGTCTGA CAGACCTTCG CAGTCTCATG	7620
GGCTATGTTT CTCAGGACCA GTTCTTTTTT GCGACTTCAA TCCTAGACAA TATCCGCTTT	7680
GGCAATCCTA ACTTGCCCCCT TTCAGCGGTC GAGGAAGCTA CTAAGCTAGC CCGGGTTTAC	7740
CAAGATATTG TAGACATGCC TCAAGGATTT GATACGCTGA TTGGTGAAAA AGGAGTCACT	7800
CTTTCTGGTG GTCAAAAGCA ACGGTTGGCT ATGAGTCGGG CTATGATTTT AGACCCTGAT	7860
ATCTTGATTT TGGATGATTC CTTATCCGCC GTAGATGCCA AGACAGAGTA TGCGATTATC	7920
GACAACCTCA AGGAGATGCG AAAGGACAAG ACAACCATTA TCACTGCCCA TCGCCTCAGT	7980
GCTGTTGTCC ATGCAGATTT TATTTTAGTT CTACAAAATG GTCAAATTAT CGAACGAGGC	8040
ACGCACGAAG ACTTGCTAGC TTTGGATGGC TGGTATGCCC AAACCTACCA GTCTCAGCAG	8100
TTGGAAATGA AAGGAGAAGA AGATGCAGAA TAAACAAGAA CAATGGACTG TATTGAAGCG	8160
CTTGATGTCT TATCTCAAGC CTTATGGACT CCTGACCTTT TTGGCACTCA GTTTTCTCCT	8220

572

AGCGACGACG GTCATTAAAA GTGTCATACC CCTCGTGGCT TCCCACTTTA TCGACCAGTA	8280
TCTCAGCAAT CTTAACCAAC TAGCCGTTAC CGTTTGGCTG GTCTACTATG GTCTCTACAT	8340
CCTACAAACT GTAGTTCAGT ATGTCGGCAA TCTTCTCTTT GCGCGCGTGT CTTACAGTAT	8400
TGTTAGGGAT ATTCGTGGG ATGCCCTTGC CAATATGGAG AAACGGGCA TGTCTTACTT	8460
TGACAAGACG CCAGCAGGTT CTATCGTTTC TCGTTTGACC AACGATACCG AGACGATTAG	8520
TGATATGTTT TCTGGGATTT TATCCAGCTT TATCTCAGCA GTTTTTATCT TTCTGACAAC	8580
CCTTTATACC ATGTTGGTGC TGGATTTTCG TTGACGGCT TTAGTCTTGC TCTTCTTCC	8640
TTTGATTTTC CTTTGGTCA ATCTCTATCG AAAAAAGTCA GTGAAAATCA TCGAGAAAAC	8700
CAGAAGTCTC TTGTCAGATA TCAATAGTAA GCTGGCAGAG AATATCGAGG GAATCAGGAT	8760
TATTCAGGCC TTAAATCAAG AGAAGCGCCT GCAGGCAGAA TTTGATGAAA TCAACCAAGA	8820
ACACTTGGTC TACGCCAACC GTTCTGTAGC CTGGATGCC CTCTTTTGA GACCTGCCAT	8880
GAGTTTGGCTG AAACCTCTAG GCTATGCAGT CTGATGGCC TACTTTGGCT ACCGTGGTTT	8940
TTCTATCGGG ATAACGGTCG GGACCATGTA TGCCTTTATC CAGTACATCA ACCGCCTTTT	9000
TGACCCCTTG ATGAGGTGA CGCAAACTT TTCAACTCTG CAAACGGCTA TGGTTTCTGC	9060
AGGTCGTGTC TTGCCCCTGA TAGACGAGAG GACCTATGAA CCTCTTCAAG AAAATGGGCA	9120
AGCCAAAGTC CAAGAAGGCA ATATCCGTTT TGAACATGTG TGTTCCTCAT ATGACGGTAA	9180
ACATCCGATT CTGGATGACA TTTCTTTCTC TGTTAATAAG GGTGAAACCA TTGCCTTTGT	9240
AGGTCATACA GGTTCAGGGA AATCGTCTAT TATCAATGTC CTCATGCGCT TTTATGAATT	9300
CCAGTCAGGG AGAGTTCTCT TGGATGATGT GGATATCAGG GATTTTCAGTC AAGAAGAGCT	9360
GAGAAAAAAC ATCGGTTTGG TCTTGCAGGA ACCCTTCCTC TATCATGGAA CTATTAAATC	9420
CAATATCGCC ATGTACCAAG AAACCAAGTGA TGAGCAGGTT CAGGCTGCGG CAGCCTTTGT	9480
GGATGCAGAT TCCTTTATTC AAGAACTTCC TCAGGGGTAC GACTCCCCTG TTTCCGAGCG	9540
TGGTTCGAGC TTCTCTACTG GGCAACGCCA GCTTCTTGCC TTGCTAGAA CAGTCGCCAG	9600
CCAGCCTAAA ATCCTGATTT TCGATGAAGC GACAGCCAAT ATTGACTCTG AAACAGAAAG	9660
CTTGGTTCAA GCTTCTCTGG CGAAGATGAG ACAGGGCCGA ACAACTATTG CTATCGCTCA	9720
CCGCCTTTCT ACTATTCAAG ATGCCAACTG CATCTATGTC TTGGATAAGG GACGCATTAT	9780
CGAGAGTGGA ACCCATCAGG AACTCTTGCC TCTGGGAGGA ACCTATCACA AGATGTATAG	9840
TTTGCAGGCA GGGGCCATGG CCGATACTCT TTGAAAATCT CTTTAAACCA TGTACGCTTT	9900
ATCTGCAATC TCAAAGCTGT ACTTTGATTT TCATTGAGTA CTAGAAGGAA ATCCTTCAAA	9960
TTACAGATTT CTTTCACCGC CTTTTCATT TTCTGGTATA ATGAAAAATG TTGACAAATA	10020

573

GTATAATAAA AACAAAGGAG AACAGCATGC TGAAATGGGA AGACTTGCCT GTGGAAATGA	10080
AATCAAGCGA GGTGAGTCT TACTACCAGC TTGTCTCTAA AAGGAAGGGT TCGCTGATTT	10140
TCAAGCGTTG CTGGACTGG GTTTTGGCCT TGGTCTTACT GGTCTGACC TCTCCCATCT	10200
TTCTCATCTT GAGCATTGG ATCAAGTTGG ATAGCAAAGG GCCAGTGATT TACAAGCAAG	10260
AGCGTGTGAC CCAGTACAAC CGTCGGTTCA AGATTTGGAA GTTTCGTACC ATGGTGACGG	10320
ATGCGGATAA AAAAGGAAGT CTGGTGACTT CTGCTAACGA TAGCCGCATT ACCAAGGTTG	10380
GAAATTTTCAT CCGACGTGTC CGTTTGGACG AACTGCCTCA GTTGGTCAAT GTCCTTAAAG	10440
GTGAGATGTC CTTTGTCSGT ACACGACCTG AAGTGCCACG TTATACAGAG CAGTATAGCC	10500
CTGAAATGAT GGCAACCTTG CTCTTGCAAG CAGGGATTAC CTCTCCAGCC AGCATCAACT	10560
ACAAGGATGA GGACACAATT ATCAGTCAAA TGACGGAGAA AGGTCTGTCA GTTGATCAGG	10620
CCTATGTGGA GCATGTTCTT CCTGAAAAGA TGCGCTATAA CCTCGCCTAT CTCCGAGAGT	10680
TTAGTTTCTT TGGGGACATC AAAATCATGT TTCAAACCGT GTTTGAGGTA CTAAAATAAA	10740
GTAGTCATAA GAAAATGAGT ACAGATAAAA GGAGCAAATC AATGCCAAAT TACAATATTC	10800
CATTTTCACC GCCTGATATC ACAGAAGCAG AAATTACTGA AGTAGTGGAT ACCCTGCGTT	10860
CTGGTTGGAT CACAACAGGT CCTAAAACAA AAGAACTGGA GCGCCGCTTG TCTCTTTACA	10920
CACAGACACC TAAGACTGTT TGTCTCAACT CTGCGACAGC CGCTCTGGAG TTGATTTTAC	10980
GCGTTTTGGA AGTGGGACCT GGTGATGAAG TCATCGTTCC AGCCATGACC TATACGGCTT	11040
CATGTAGTGT CATTACGCAC GTGGGAGCAA CCCCTGTCAT GGTGGATATC CAAGCAGATA	11100
CGTTTGAGAT GGACTATGAC CTGCTTGACC AAGCTATCAC TGAGAAAAC T AAGGTGATTA	11160
TTCCAGTAGA GCTCGCAGGG ATTGTTTGCG ATTATGACCG TTTGTTCCAA GTCGTGGAGA	11220
AAAAACGTGA CTTCTTTACC GCTTCAAGCA AGTGGCAAAA GGCCTTTAAC CGTATTGTCA	11280
TTGTCTCTGA TAGTGCCAC GCTTTGGGAT CTATTTATAA AGGACAACCT TCTGGTTCTA	11340
TCCGTGACTT TACTTCCTTC TCATTCCATG CAGTTAAGAA CTTTACAACG GCAGAAGGTG	11400
GAAGTGCGAC TTGGAAAGCC AATCCAGTGA TTGATGACGA AGAGATGTAC AAGGAATTCC	11460
AAATCCTTTC CTTTCACGGG CAAACTAAGG ATGCTCTTGC CAAGATGCAA CTGGGGTCAT	11520
GGGAATACGA TATCGTTACA CCAGCCTATA AGTGCAACAT GACCGATATC ATGGCTTCAC	11580
TTGGTTTGGT ACAATTGGAC CGCTATCCAA GTTTGTTGCA ACGCCGTAAG GACATTGTGG	11640
ACCGCTATGA TAGTGGTTTT GCAGGTTCTC GCATCCATCC TTTGGCACAC AAGACTGAAA	11700
CTGTGGAATC TTCACGCCAC CTCTACATCA CCCGTGTAGA AGGAGCAAGC CTAGAAGAAC	11760

574

GCAACCTCAT CATCCAAGAA TTGGCTAAAG CAGGAATTGC AAGTAATGTT CACTACAAAC	11820
CGCTTCCTCT CTTGACAGCC TATAAGAATC TTGGATTGGA TATGACGAAC TATCCTAAGG	11890
CCTATGCCTT CTTTGAGAAT GAAATTACCC TCCCTCTTCA TACTAAATTA AGCGATGAAG	11940
AAGTAGACTA TATCATTGAG ACTTTCAAAA CAGTTTCTGA AAAAGTGCTA ACTTTATCAA	12000
AAAAATGACA AACTACAGTC AAGCGAAAGT GATCCTGCCC CTAAAAAGTC TAATTGAGTG	12060
TAAAAACTGT TGTTTCAAT TGATAATAGT TTACACCTGT AGTTGAGGCC CCTTCTCCT	12120
CAGAGAGAGA ATTTTATAG GATTTTCCTT TCTTGTTGGA GTCCCGTGGT TTGAAATAAG	12180
ATGTGAGCAA TTTAGTGTAG CATTTAGAAT CCTTACTAGA CATCATTTAG AAAATCTAGT	12240
GTCTTGTTCT AGTTTCAAT TCACCCTATT TTTTGAAAGA CGTGAGTTT CATGAGTGAG	12300
ATTGTGGAAT CTCGCGTCTT TTTTGTGTTT CAGAATATTG TTCAAATTT TGTGCTGTC	12360
TTTCATGTTT TAGTCATTCT TTTGCATGAT AGAATTTATA GCATGTTGAT ATTATAATAA	12420
TACAAATATT CTATATGTTT AGTGATGCTT GCTATACATT ATTAGATCTC CTGCGAGACA	12480
ATCTATAAAA CACTTGTCTA CGATTACCTA TATGCCCTAT TCCAGTATTT TAGAAGCACT	12540
GCATCTATTT TTATCGAGGT TAAATCTAGC TTTTATAGAA GGTCTATTTA AGAAATATAT	12600
TGTAGTGTTT TAGTTTCAAT CCGCCATATG AGCGATATTC AGGTAAATAT CCCTGGCGAA	12660
TGCTTGTATG ACAAGGTATT TGTTCTTTCA TTTATAATTT ACAACATATC AACAAATTTA	12720
AATATAGTAA ATGGGATATT TTATATTCAA GCTAAGAAAG ATAGCATCAC TTTTGAATGG	12780
AAGGCTAAAG AGCAAAGT AGAAGTTGGCC ATAGATAGCT CAAAACCTG CTTTGAGGTT	12840
GTAGATATAG TAAAATGAAA TGAGAATAGG ACAAATTGAT CGGGACAGTC AAATCGATTT	12900
CTAACAATGT TTTAGAAGTA GAGGTGTACT ATTTTAGTTT CAGTCTACTA TAGAACTGAC	12960
CAAGTCAGTA ACCTAGACTT AGGGCAAGGC GGCCTGACC TAGTTTGAAG AGATTTCGA	13020
AGAGTATAAA TTTTAATATT TTCTTGTTT ATTCCTTGAC AATTCAATTT GGAAAATATA	13080
TGATAAAGAT AATGACAGCG GTGTCAATCT ATCTATTTTA AGAAAAGTAA TAATCAATTG	13140
TTAAAAATAG TAAAAAATTT GGAGGTTCTG ATGAAATATT TTGTTCCG	13188

(2) INFORMATION FOR SEQ ID NO: 71:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32768 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

575

AACGAGTGCA TCAGTCTCAG CAAGCACCAG TCGGTCGGCC TCAGCAAGCA CCAGCGCGTC	60
TGAATCCGCA TCAACCAGTG CCTCAGCTTC AGCAAGTACC TCAGCATCTG AATCAGCATC	120
AACAAGTGCA TCGGCTTCAG CAAGCACAAG TGCTTCAGCC TCAGCAAGTA TCTCAGCGTC	180
TGAATCCGCA TCAACGAGTG CGTCCGCTTC AGCAAGTACT AGCGCCTCAG CATCAGCGTC	240
AACAAGTGCT TCGGCTTCAG CGTCAACGAG TGCCTCTGAG TCAGCATCAA CGAGTACGTC	300
AGCCTCAGCA AGCACATCAG CTTCTGAATC TGCATCAACC AGTGCGTCAG CCTCAGCATC	360
GACAAGCGCC TCAGCTTCAG CAAGTACCAG TCGCTCAGCC TCAGCAAGTA CCAGTGCTTC	420
AGCCTCAGCG TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG	480
TGCGTCAGCC TCAGCAAGTA CTAGCGCCTC AGCCTCAGCA TCAACGAGTG CGTCCGCTTC	540
AGCAAGTACT AGTGATCAG CATCAGCATC AACGAGTGCA TCGGCTTCAG CAAGTACCAG	600
CGCCTCAGCT TCAGCAAGCA CCAGTGCGTC AGCCTCAGCA AGTACCAGCG CCTCAGCCTC	660
AGCAAGCACC ACTGCCTCAG CTTCAGCAAG TACCAGTGGC TCAGCCTCAG CGTCGACAAG	720
TGCGTCGGCT TCAGCAAGTA CCTCAGCGTC TGAATCAGCA TCAACGAGTG CATCAGCTTC	780
AGCATCAACA ACTGCTTCAG CTTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG	840
TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC AGCATCAGCG TCAACAAGTG CTTCCGCTTC	900
AGCGTCAACG ACTGCGTCTG AGTCAGCATC AACGAGTACG TCAGCCTCAG CAAGCACATC	960
AGCTTCTGAA TCTGCATCAA CCAGTGCGTC AGCCTCAGCA TCGACAAGCG CCTCAGCTTC	1020
AGCAAGTACC ACTGCGTCAG CCTCAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG	1080
TGCGTCGGCC TCAACCAGTG CATCTGAATC GGCATCAACC AGTGCGTCAG CCTCAGCAAG	1140
TACTAGCGCC TCAGCCTCAG CATCAACGAG TSCGTCCGCT TCAGCAAGTA CTAGTGATC	1200
AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC AGCGCCTCAG CTTCAGCAAG	1260
CACCAGTGGC TCAGCTCAG CAAGTACCAG CGCCTCAGCC TCAGCAAGCA CCAGTGCCCTC	1320
AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG	1380
TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT TCAGCATCAA CAAGTGCTTC	1440
AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG AGTGCTTCAG TCTCAGCGTC	1500
AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT TCAGCAAGCA CCAGTGCGTC	1560
GGCTTCAGCA AGTACTAGTG CATCGGCTTC AGCATCGACA AGTGCGTCTG AATCGGCATC	1620
AACGAGTGCT TCGGCTTCAG CATCAACGAG TCGGTCAGCC TCAGCAAGCA CATCAGCTTC	1680
TGAATCTGCA TCAACCAGTG CGTCCGCTTC AGCGTCAACC AGTGCGTCGG CTTCAGCGTC	1740

576

GACAAGTGCT TCGGCTTCAG CATCAACGAG TCGGTCGGCC TCAGCAAGCG CAAGTACCTC	1800
AGCGTCAGct TCCGCCTCAA CCAGTGCGTC GGCTTCAGCA AGCACAAGTG CGTCAGCCTC	1860
AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCTGAGTCAG CATCAACGAG	1920
TACGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAGTG CGTCAGCCTC	1980
AGCATCGACA AGCGCCTCAG CTTAGCAAG TACCAGTGCT TCAGCCTCAG CGTCGACAAG	2040
TGCGTCGGCC TCAACCAGTG CATCTGAATC GGATCAACC AGTGCGTCAG CCTCAGCAAG	2100
TACTAGTGCA TCAGCTTCAG CATCAACGAG TGCATCGGCT TCAGCATCAA CCAGTGCCCTC	2160
GGCTTCAGCG TCAACCAGTG CGTCAGCTTC AGCAAGTACC AGTGCTTCAG TCTCAGCATC	2220
AACAAGTGCT TCAGCCTCAG CATCGACAAG TGCCTCGGCT TCAGCAAGCA CATCAGCATC	2280
TGAATCAGCG TCAACCAGTG CTTGGGCTTC AGCAAGTACC AGTGCTTCAG CTTCAGCATC	2340
AACCAGCGCC TCGGCCTCAG CAAGCACCTC AGCTTCTGAA TCGGCCTCAA CCAGCGCCTC	2400
GGCTTCAGCA AGCACCTCAG CTTCTGAATC GGCTCAACC AGCGCCTCAG CCTCAGCATC	2460
AACGAGTGCT TCGGCTTCAG CAAGCACAAG CGCCTCGGGT TCAGCATCAA CGAGTACGTC	2520
AGCTTCAGCG TCAACCAGTG CTTAGCCTC AGCATCAACA AGTGCGTCAG CCTCAGCAAG	2580
TATCTCAGCG TCTGAATCGG CATCAACGAG TCGCTCTGAG TCAGCATCAA CGAGTACGTC	2640
AGCCTCAGCA AGCACCTCAG CTTCTGAATC GGCTCAACC AGTGCGTCAG CCTCAGCATC	2700
GACAAGCGCC TCAGCTTCAG CAAGTACCAG TGCTTCAGCC TCAGCGTCGA CAAGTGCGTC	2760
GGCTCAACC AGTGCACTG AATCGGCATC AACCAGTGCG TCAGCCTCAG CAAGTACTAG	2820
TGCATCGGCT TCAGCATCAA CCAGTGCCCTC GGCTTCAGCG TCAACCAGTG CGTCAGCTTC	2880
AGCAAGTACC AGTGCTTCAG TCTCAGCATC AACAAGTGCT TCAGCCTCAG CATCGACAAG	2940
TGCCTCGGCT TCAGCAAGCA CATCAGCATC TGAATCAGCG TCGACAAGCG CCTCAGCTTC	3000
AGCAAGTACC AGTGCGTCAG CCTCAGCGTC GACAAGTGCG TCAGCCTCAG CAAGTACTAG	3060
TGCATCAGCT TCAGCATCAA CGAGTGCATC GGCTTCGGCG TCAACCAGTG CATCAGAGTC	3120
AGCAAGTACC AGTGCGTCag CTTCCGCATC AACAAGTGCC TCGGCTTCAG CAAGCACCAG	3180
TGCGTCGGCT TCAGCAAGTA CTAGCGCCTC AGCCTCAGCC TCAACCAGTG CGTCAGCCTC	3240
AGCAAGTATC TCAGCGTCTG AATCGGCATC AACGAGTGCG TCCGCTTCAG CAAGTACTAG	3300
CGCCTCAGCC TCAGCGTCAA CAAGTGCATC GGCTTCAGCG TCAACGAGTG CGTCTGAATC	3360
GGCATCAACG AGTGCGTCCG CTTAGCAAG TACTAGCGCC TCAGCCTCAG CGTCAACAAG	3420
TGCATCGGCT TCAGCATCAA CGAGTGCCTC CGCTTCAGCA AGTACTAGCG CCTCAGCCTC	3480
AGCGTCAACA AGTGCACTCG CTTAGCGTC AACGAGTGCG TCTGAGTCAG CATCAACGAG	3540

577

TGCGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA TCAACCAGTG CGTCAGCCTC	3600
AGCATCGACA AGCGCCTCAG CTTAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG	3660
TGCGTCGGCT TCAGCAAGTA CCAGTGCGTC AGCCTCAGCA AGTACCAGTG CGTCAGCCTC	3720
AGCGTCGACA AGTGCGTCGG CCTCAACCAG TGCATCTGAA TCGGCATCAA CCAGTGCGTC	3780
AGCCTCAGCA AGTACTAGTG CATCAGCTTC AGCATCAACG AGTGCATCGG CTTAGCATC	3840
AACCAGTGCA TCAGAGTCAG CAAGTACCAG TGCGTCAGCT TCCGCATCAA CAAGTGCCTC	3900
GGCTTCAGCA AGTACTAGCG CCTCAGCCTC AGCGTCAACA AGTGCTTCAG CTTCCGCGTC	3960
AACCAGCGCC TCGGCCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CAAGTGCCTC	4020
GGCTTCAGCA TCAACGAGTG CATCAGTCTC AGCAAGCACC AGTGCGTCGG CCTCAGCAAG	4080
CACCAGCGCG TCTGAATCCG CATCAACCAG TGCCTCAGCT TCAGCAAGTA CCTCAGCATC	4140
TGAATCAGCA TCAACAAGTG CCTCGGCTTC AGCAAGCACA AGTGCTTCAG CCTCAGCAAG	4200
TATCTCAGCG TCTGAATCGG CATCAACGAG TGCGTCCGCT TCAGCAAGTA CTAGCGCCTC	4260
AGCATCAGCG TCAACAAGTG CTTCCGCTTC AGCGTCAACG AGTGCGTCTG AGTCAGCATC	4320
AACGAGTACG TCAGCCTCAG CAAGCACATC AGCTTCTGAA TCTGCATCAA CCAGTGCGTC	4380
AGCCTCAGCA TCGACAAGCG CCTCAGCTTC AGCAAGTACC AGTGCGTCAG CCTCAGCAAG	4440
TACCAGTGCT TCAGCCTCAG CGTCGACAAG TGCGTCGGCC TCAACCAGTG CATCTGAATC	4500
GGCATCAACC AGTGCGTCAG CCTCAGCAAG TACTAGCGCC TCAGCCTCAG CATCAACGAG	4560
TGCGTCCGCT TCAGCAAGTA CTAGTGATC AGCTTCAGCA AGTACTAGCG CCTCAGCCTC	4620
AGCGTCGACA AGCGCCTCAG CTTAGCAAG TACCAGTGCG TCAGCCTCAG CGTCGACAAG	4680
TGCGTCGGCT TCAGCAAGTA CCTCAGCCTC TGAATCAGCA TCAACAAGTG CGTCGGCTTC	4740
AGCATCAACG AGTGCATCAG CTTAGCATC AACAAAGTGCT TCAGCTTCAG CAAGTACCAG	4800
TGCGTCGGCT TCAGCATCAA CGAGTGCTTC AGTCTCAGCG TCAACCAGTG CCTCTGAATC	4860
CGCATCAACA AGTGCCCTCG CTTAGCAAG CACCAGTGCT TCGGCTTCAG CGTCAACGAG	4920
TGCGTCTGAG TCAGCATCAA CGAGTGCGTC AGCCTCAGCA AGCACATCAG CTTCTGAATC	4980
TGCATCAACC AGTGCGTCAG CTTCCGCATC AACAAAGCGCC TCGGCCTCAG CAAGTACAAG	5040
TGCTTCAGCC TCAGCATCAA CCAGTGATC AGCTTCAGCC TCAACAAGTG CTTAGCCTC	5100
AGCGTCAACC AGTGCCCTCG CTTAGCAAG TACCAGTGCG TCAGCTTCAG CAAGCACAAG	5160
TGCGTCAGCT TCAGCATCAA CCAGTGCTTC GGCTTCGGCA TCAACAAGTG CCTCAGCATC	5220
AGCATCAACG AGTGCGTCAG CCTCAGCAAG TACTAGTGCA TCAGCATCAG CATCAACCAG	5280

578  
TGCATCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA TCAACGAGTG CATCAGCATC 5340  
AGCATCAACG AGTGCATCGG CTTCAGCGTC AACCAGTGCA TCAGTCTCAG CAAGCACCAG 5400  
TGGCTCGGCT TCAGCATCAA CGAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC 5460  
GGCATCAACG AGTGGCTCAG CCTCAGCAAG TACTAGTGCA TCGGCTTCAG CAAGCACCAG 5520  
TGGCTCGGCT TCAGCATCAA CCAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC 5580  
GGCATCAACG AGTGGCTCAG CCTCAGCAAG TACTAGTGCA TCAGCmTCAG CATCAACGAG 5640  
TGCATCGGCT TCAGCAAGTA CCAGCGCCTC AGCTTCAGCA AGCACCAGTG CGTCAGCCTC 5700  
AGCAAGTACC AGCGCCTCAG CCTCAGCAAG CACCAGTGCC TCAGCTTCAG CAAGTACCAG 5760  
TGGCTCAGcT CAGCATCAAC AAGTGCTTCA GCTTCGGCCT CAACAAGTGC GTCAGCTTCA 5820  
GCATCAACGA GTGGCTCGGC TTCAGCAAGC ACCAGTGCCCT CGGCCTCAGC AAGCACCAGT 5880  
GCTTCAGCTT CAGCATCAAC AAGTGCGTCA GCTTCAGCAA GTACATCAGT TTCAAATTCA 5940  
GCAAACCATT CGAACTCACA AGTTGGAAAT ACTTCTGGAT CGACAGGTAA ATCCCAAAAA 6000  
GAATTGCCTA ATACAGGTAC TGAGTCGTCA ATTGGATCTG TGTACTTGG AGTCTAGCA 6060  
GCTGTTACAG GTATTGGATT GGTTCGGAAC CGCCGTAAAC GTGATGAAGA AGAGTAAGAC 6120  
AACCTGTAAA GTTAGGCTAA ACTAACTCGC GCACATAAAT CAAGGAGAAA ATTGCTAGTG 6180  
GATGATAAAA TAACAGTCAT TGTACCAGTA TACAATGTGG AAAACTATCT GAGGAAGTGC 6240  
CTAGATAGTA TTATTACTCA AACATATAAA AATATTGAGA TTGTTGTCCT TAATGATGGT 6300  
TCTACGGATG CTTCAGGTGA AATTTGTAAA GAATTTTCAG AAATGGATCA CCGAATTCTC 6360  
TATATAGAAC AAGAAAATGC TGGTCTTTCT GCCGCACGAA ACACCGGTCT GAATAATATG 6420  
TCCGGAAATT ATGTGACCTT TGTGGACTCG GATGATTGGA TTGAGCAAGA TTATGTAGAA 6480  
ACTCTATATA AAAAAATAGT ACAGTATCAG GCTGATATTG CAGTTGGTAA TTATTATTCT 6540  
TTCAACGAAA GTGAAGGAAT GTTCTACTTT CATATATTGG GAGACTCCTA TTATGAGAAA 6600  
GTATATGATA ATGTTTCTAT CTTTGAGAAC TTGTATGAAA CTCAAGAAAT GAAGAGTTTT 6660  
GCTTTGATAT CTGCTTGGGG TAAACTCTAT AAGGCAAGAT TGTTTGAGCA GTTGCGCTTT 5720  
GACATAGGTA AATTAGGAGA AGATGGTTAC CTCAATCAAA AGGTATATTT ATTATCAGAA 6780  
AAGGTAATTT ATTTAAATAA AAGTCTTTAT GCTTATCGGA TTAGAAAAGG TAGTTTATCA 6840  
AGAGTTTGGA CAGAAAAGTG GATGCACGCT TTAGTTGATG CTATGTCTGA ACGTATTACG 6900  
CTACTAGCTA ATATGGGTTA TCCTCTAGAG AAACACTTGG CAGTTTATCG TCAGATGTTG 6960  
GAAGTCAGTC TCGCCAACGG TCAAGCTAGT GGTttTATCTG ACACAGCAAC GTATAAAGAG 7020  
TTTGAAATGA AACAAAGGCT TTAAATCAG CTATCGAGAC AAGAGGAAAG TGAAAAGAAA 7080



579

GCCATTGTCC TCGCAGCAAA CTATGGCTAT GTAGACCAAG TTTTAACGAC AATCAAGTCT	7140
ATTTGTTATC ATAATCGTTC GATTTCGTTT TATCTGATTC ATAGCGATTT TCCAAATGAA	7200
TGGATTAAGC AATTAAATAA GCGCTTAGAG AAGTTTGACT CAGAAATTAT TAATTGTCGG	7260
GTAACCTCTG AGCAAATTC ATGTTATAAA TCGGATATTA GTTACACAGT CTTTTACGC	7320
TATTTCATAG CTGATTTCTG GCAAGAAGAC AAGGCCCTCT ACTTGGACTG TGATCTAGTT	7380
GTAACGAAAA ATCTGGATGA CTTGTTTGCT ACAGACTTAC AAGATTATCC TTTGGCTGCT	7440
GTTAGAGATT TTGGGGGCAG AGCTTATTTT GGTCAAGAAA TCTTTAATGC CCGTGTCTC	7500
TTGGTAAACA ATGCTTTTGG GAAAAAGAG AATATGACCC AAAAATTAAT TGATGTAACC	7560
AATGAATGGC ATGATAAGGT GGATCAGGCA GATCAGAGCA TCTTGAATAT GCTTTTGAA	7620
CATAAATGGT TGGAATTGGA CTTTGATTAT AATCATATTG TCATTCATAA ACAGTTTGCT	7680
GATTATCAAT TGCCTGAGGG TCAGGATTAT CCTGCTATTA TTCACTATCT TTCTCATCGG	7740
AAACCGTGGA AAGATTGGC GGGCCAAACC TATCGTGAAG TTTGGTGGTA CTATCATGGG	7800
CTTGAATGGA CAGAATTGGG ACAAACCAT CATTTACATC CATTACAAAG ATCTCACATC	7860
TATCCAATAA AGGAACCTTT CACTTGTCTA ATCTATACTG CCTCAGACCA TATTGAACAA	7920
ATTGAGACAT TGGTTCAATC CTTGCCTGAT ATTCAGTTTA AGATAGCAGC TAGAGTAATA	7980
GTTAGTGATC GATTGGCTCA GATGACAATT TATCCAAACG TGACTATATT TAACGGAATT	8040
CACTATTTGG TAGATGTCGA TAATGAATTG GTAGAAACCA GTCAAGTACT TTTAGATATT	8100
AATCATGGCG AAAAGACAGA AGAAATTCTC GATCAATTTG CTAATCTTGG CAAGCCTATC	8160
TTATCCTTTG AAAATACTAA AACCTATGAA GTAGGTCAGG AGGCATATGC TGTTGACCAA	8220
GTTCAAGCAA TGATTGAAAA ATTGAGAGAA ATAAGCAAAT GAAGAAAAAT CATTTAGTAG	8280
GAGATGCTCT GATTTTGACC GTTACTGATC AGATTGAAGA GTTGGATTAT TTTTATAAAA	8340
ATTTCTCCGT TCATCATATA TGAAAGTTCT TCAAACATCA GAGTGCTTTA TAAATATAA	8400
ATAGACCTAA AGATATTTAA TATGAACTGC ACCCCAAAAG TTAGACAGAA AAAATCTAAC	8460
TTTTTGGsGT CAGTACAATA TTAGGGTGTG ATTAATTATC TTTTAGGTG AAAATGATTC	8520
TATATTATAG CTGTTTGATA CGAAATTTAT TATAAGGAAA TTATGTTAAT GAATACAAAA	8580
TCTATAGTTT TTAATGCAGA TAATGATTAT GTAGATAAAT TAGAACTGC AATTAAATCT	8640
ATTGTTGTT ATAATAATTG TTTAAAATTT TATGTATTTA ATGATGATAT TGCGTCAGAG	8700
TGTTTTTTGA TGATGAATAA GCGATTGAAG ACTATACAAT CTGAAATCGT TAATGTAAAG	8760
ATTGTAGATC ATGTTCTTAA AAAGTTTCAT TTACCGTTAA AGAATTTAAG TTATGCCACT	8820

580  
TTCTTTCGTT ATTTTATACC TAATTTTGTG AAAGAAAGTC GTGCTTTATA CCTAGATTCT 8880  
GACATCATTG TTACAGGAAG TTTAGACTAT TTATTTGATA TAGAACTAGA TGGTTATGCC 9940  
TTGGCAGCAG TAGAAGATTC TTTTGGTGAT GTTCCTTCTA CCAATTTTAA CTCGGAATG 9000  
TTATTAGTTA ATGTAGATAC TTGGAGAGAT GAAGATGCTT GTTCGAAACT GTTAGAACTG 9060  
ACCAATCAAT ATCATGAAAC AGCATATGGA GATCAAGGAA TTTTAAATAT GTTATTCCAT 9120  
GATAGATGGA AAAGATTAGA CCGAAATTTT AATTTTATGG TGGGGATGGA TAGCGTCGCA 9180  
CACATAGAAG GAAATCATAA ATGGTATGAG ATTCTGAGT TGAAAAATGG AGATTACCT 9240  
AGTGTATAC ATTATACTGG GGTAAACCT TGGGAAATAA TTTCCAATAA TCGCTTTAGA 9300  
GAAGTTTGGT GGTTTTATAA TCTGTTAGAA TGGTCTGATA TTTTATTGAG AAAAGACATT 9360  
ATTAGTCGTA GTTTCGAAGA ACTTGATAC AGTCCTAAAG CTCATACAGC AATTTTACA 9420  
GCTAGTTGTG AGATGGAGCA TGTAGAATAT TTGATAGAAA ATTTACCAGA GGTACATTTT 9480  
TCTATACTAG CACATACATA TTTTGGCTCT AGTGTGCTT CTTTATTAAG ATATAGCAAT 9540  
GTTACGATTT ATCCTTGTCT TTCTCCATTT GATTATCGAA AAATTTTGA TAATTAGAT 9600  
TTTTATTTAG ATATTAATCA TTATAAGAA GTGGATAATA TTGTATCCGT TGTTCACAA 9660  
CTATCTAAAC CAATTTTAC CTTTGAAT ACTAGTCATG ATATAGGCAA TCAAACATA 9720  
ATATTTTCTT CAACCGAACC AAACAAATG GTAGAGGCTA TTAGACAATT TATAGGAGAA 9780  
TAAGTTTATG GCAGACGAAC TAATTAGTAT TGTAGTTCCA ATCTACAACG TTGAGAATTA 9840  
TTTGGCAATG TGTGTGATA GCATTCAGAA TCAGACGTAT CAAAATTTT AGTGTTTATT 9900  
AATCAATGAT GGCTCTCCAG ATCATTTCAT CAAAATATGT GAAGAATTTG TAGAGAAACA 9960  
TTCTCGTTTC AAATATTTT AGAAAGCAAA CGCGGTCTT TCATCAGCTC GTAACCTAGG 10020  
TATTGAATGT TCGGGGGGGG CGGTACATTA CTTTGTAGA CTCTGATGAT TGGTTGGAAC 10080  
ATGATGCTTT AGACCGATTA TATGGTGCTT TGAAGGGA AAACGCAGAT ATTAGTATCG 10140  
GGCGTTATAA TTCTTATGAT GAAACACGCT ATGTGTATAT GACTTATGTT ACGGATCCAG 10200  
ATGATTCTCT AGAAGTGATA GAAGGTAAAG CAATTATGGA TAGGGAAGGT GTCGAAGAAG 10260  
TCAGAAATGG GAACTGGACT GTAGCTGTCT TGAAGTTATT CAAGAGAGAG TTAATAACAG 10320  
ATTTACCATT TCCTATAGGA AAAATTGCAG AGGATACTTA CTGGACATGG AAGGTACTTC 10380  
TAAGAGCTTC GAGGATAGTC TATTTGAATC GTTGTGTTTA CTGGTACCGT GTTGGTTTAT 10440  
CTGATACTTT ATCGAATACA TGGAGTGAAA AGCGTATGTA TGATGAAATT GGGGCTAGGG 10500  
AAGAAAAGAT AGCTATTTTA GCAAGTTCAG ACTATGACTT GACCAATCAT ATTTTGATTT 10560  
ATAAAAATAG ATTACAAAGA GTGATAGCAA AATTAGAAGA ACAAATATG CAGTTCACAG 10620

581

AGATTTACAG AAGAATGATG GAAAAATTGT CTTTACTTCC GTAGATAGTA ATAAAAAATG	10680
AGATAGCGTA ATATGAAACT ACATTTAACA AATTTATACG GCATGCCTGG TGATAGTACG	10740
GTTATCTTAG CTCAAAATGC TGTTCAAAAG ATAGCTAGTC AACTGGGATT TAGAGAGGTT	10800
GGTATTTATT TTTACAACAT TGCTTCAGAT AGTCCTTCTG AAATGAATAA GCGTCTGGAT	10860
GGTATTATGG CCAGTATCTC TATTGGGGAT ATTTTAGTCT TTCAGTCTCC AACCTGGAAT	10920
GGTTTTGAAT TTGATCGTCT CTTGTTTGAT AAGCTAAAGG ATATGCAGGT GAAAATTATT	10980
TGCTTTATCC ATGATGTTGT TCCCCTCATG TTGATAGTA ACTATTATCT CATGAAAGAT	11040
TATCTGTATA TGTATAATCT ATCAGATGTT TTGATAGTGC CCTCAGAGAG AATGAAAACA	11100
CGCCTGATGG AAGAAGGATT GACGACTAAG AAGATTCTTG TTCAAGGGAT GTGGGATCAT	11160
CCTCATGATT TATCCTTATA CACCCCTGCT TTTAAAAAAG AACTTTTTTT TGCTGGAAGT	11220
TTAGAGCGTT TTCCAGACTT ACAAATTGG TCTCAAGATA CGCCTTTGAG AGTATTTTCA	11280
AATAAAGGGG AAGCTAGTTC TAGTGCTAGA AGTCTCAGCA TCGAAGGATG GAAAAAAGAT	11340
GAGGAATTGT TGCTAGAATT ATCAAAGGGT GGATTTGGCC TTGTCTGGGG AACCCATCAA	11400
AATGAGGGAG AAAGTAACCA ATACTATACC TTGAATATAT CTCATAAGGT GAGTACCTAT	11460
CTAACAGCGG GCATTCCAGT CATTGTACCA AGTAGCTTGT CAACTGCTAA ATTTATAGTA	11520
GATCAAGGCT TGGGCTTTAT GCGCGATAGT CTGGAAGAGG TTCATGAGAT AGTTGATAAA	11580
ATGAATCTAC AAGAATATCA AGAAATGACG AATCGTATCA AGACCTTTAG CTATTTGTTA	11640
AAAGAGGGCT ATTTCACTAA AAAGTTATTG GTAGATGCAA TCTATCACTT GGAATTGAT	11700
TAAGGGAATG AAATGAACAA AACAATTGTA CTAGCAGGGG ATCGCAATTA CACCAGGCAG	11760
TTAGAAACAA CGATAAAATC TATTTTATAC CACAATCGAG ATGTTAAGAT TTATATTTTG	11820
AATCAAGATA TCATGCCAGA TTGGTTTCGC AAACCACGAA AAATAGCTCG CATGTTAGGT	11880
AGTGAGATTA TCGATGTTAA ACTACCTGAA CAACTGTCT TTCAAGATTG GGAAAAGCAA	11940
GATCACATTA GTAGCATTAC TTATGCTAGA TATTTTATTG CAGATTATAT CCAAGAAGAT	12000
AAGGTTTTAT ATTTAGACAG TGATTGATT GTAAATACTT CTTTAGAGAA ATTATTTAGT	12060
ATTTGTTTAG AAGAAAAATC ACTCGCAGCA GTTAAAGATA CAGATGGAAT TACATTTAAT	12120
GCAGGTGTTT TATTAATCAA CAATAAAAAA TGGCGTCAAG AGAAATTAAA AGAACGACTA	12180
ATTGAACAGA GCATTGTTAC AATGAAGGAA GTTGAAGAAG GCCGTTTCGA GCATTTTAAT	12240
GGTGATCAAA CGATTTTAA TCAGGTCTTG CAAGATGATT GGTTAGAACT AGGTCGAGCT	12300
TATAATTTAC AAGTAGGCCA TGATATTGTG GCTTTGTATA ACAATTGGCA GGAACATCTG	12360

582

GCTTTTAATG ATAAACCAGT GGTGATTCAT TTTACGACCT ACAGAAAACC CTGGACTACC	12420
TTGACAGCCA ATCGTTATCG TGATTATGG TGGGAATTCC ATGATTTGGA GTGGAGTCAG	12480
ATTTTACAAC ACCATATGGG AGAATTGAA CTAATATCGC CTCTAGATAA GGAATTTTCT	12540
TGCTTAACCT TAACGAATTC CCAAGATTTA GAAGGAATAG AAGAGCTAGT TACAGCTCTA	12600
CCTGAGGTGG TATTTTCATAT CGCAGCTTGG ACGGATATGG GAGATAAATT AAAAAAATTA	12660
GCTGTATATA ATAATGTGAG ATTGCATCCA CAAATGTTC CACCGGTCTT AGATAAGCTG	12720
AAAAAGTCAA CAAATCTATA TTTGGATATC AATCATGGTA GTGCAGATGA GAACTTTTTA	12780
AAATCTTTGC AAGAACAAGA AAAAACGCTA CTAGCTTTTC AATCGACTCA GCACGGAGAG	12840
TTAGGACAAA TCGTTTTCSA AAATGGGAAA GTTTCCTTTA TGATTGATAC GATTAAAGAT	12900
TTTAAGAAAA ACGGACATCT TACCTGTTT CGACAACTTC CAAGTTTAAC TTGTTTAACG	12960
TTTACGGCTT CTCAGTATAT CGAACAAATG GATTACTTGG CTGGACAGTT GCCAAATGTT	13020
GTTTTTCAA TTGCTGCTTG GACAGCTATG GGGCCAAAAT TATATGATTT GTCTAATCGT	13080
TATCCTAATA TTCAGCTCTA TCCGGCAATT TCTAGAGATA AGCTAGACGA GTTGAAGGAG	13140
AAGATGGATG CTTATTTAGA TATCAACCTA CTGACTTCAA CATCCGATAT CGTTGCAGAA	13200
ATGGCTCATC TATCTAAACC TATACTAGCC TTTTATAAAT CTCAAAATGG GAATAATGGC	13260
CAAAGGTTGT ATTCAAGTGA ACATCCTGAA CGAATGTTGG CTGATTTGCA AAAATTGATA	13320
ACTAAGGATA TGCTAGAAAA ACCGCTTGAT ATAATCCAGG TGAAAGGGAT AGATGAAACC	13380
TTGGATTATA TTATTGAACA CAACTCTTCT TTAGTTCCTT TTGGAGATGG GGAAATCAAT	13440
ATGCTTGCAG GGCATTCAAT TCCCTACCAG GATTATGATG AAGAGTTGGT TTCAATCATG	13500
AGGGACATTA TCGGCCAAGA AAGTCGAGAA GATTTAGTAG TGTGCCTTCC TGATGCTTTT	13560
ACAGATCGTT TTAGGTTTAC ATCGTGGGGG ATTCCATTTT GGAAAGATCA CATGGATCAT	13620
TATATGGATT TTTACAGAGA GTTATGCAGT GATTCATGGT ATGGCTCAAC CTTTGTATCT	13680
CGCCCTTATA TCGATTTTGA AGACAAGAGT CAAGCTAAAG CTCAATTTGA AAAATTGAAA	13740
AGCATTTGGG AAAACCGTGA CTTACTGATA GTCGAAGGTG CGACTTCTCG TTCAGGTGTC	13800
GGAAATGATT TATTGATGA GGCAAATTCT ATTAAGCGAA TTATCTGTCC TTCTCATAGT	13860
GCCTTTTCTA GAGTTCATGA ACTTGAACAA GAAATTGAAA AGTATGCTGG TGGTCGCTTG	13920
ATTTTATGTA TGCTTGACC TACAGCAAAA GTTCTGAGTT ATAATCTATG CCAGATGGGC	13980
TATCAAGTTT TGGATGTAGG CCATATTGAC TCAGAGTATG AATGGATGAA AATGGGAGCT	14040
AAAACATAAGG TTAAATTTTC TCATAAACAT ACTGCAGAAC ATAATTTTGA CCAAGATATT	14100
GAATTTATTG ATGATGAAAC CTATAACAGT CAGATTGTTG CACGAATATT AAACCTAGACT	14160

583

ATTTAAAATA AATGATAAGG ATTTAAAATG AGAAATACCA AACGCGCTGT AGTATTTGCA	14220
GGTGATTACG CTTATATTCTG ACAAATCGAA ACGGCGATGA AGTCACTCTG TAGACACAAT	14280
AGTCATTTGA AAATTTATCT GCTAAATCAG GACATTCCTC AGGAATGGTT TAGTCAAATA	14340
AGAATATATT TACAAGAGAT GGGGGGCGAC TTGATTGACT GCAAGTTAAT TGGCTCACAG	14400
TTTCAATGA ATTGGTCTAA TAAATTACCT CATATCAATC ATATGACATT TGCACGCTAT	14460
TTTATTCAG ATTGTGTAAC AGAAGATAAA GTTCTCTATC TAGATAGTGA TTTGATTGTG	14520
ACTGGTGATT TGACCGATT GTTTGAATTA GACTTAGGTG AAAATTATTT GGCAGCAGCT	14580
CGTCTTGCT TTGGAGCAGG AGTCGGCTTC AATGCTGGTG TTCTCTTGAT TAACAACAAA	14640
AAATGGGGAT CTGAAACTAT TCGACAAAAA TTGATTGACT TAACAGAAAA AGAACATGAG	14700
AATGTGGAAG AAGGAGACCA GTCAATTTTG AATATGTTGT TTAAAGATCA ATATAGTTCC	14760
CTTGAAGATC AATATAATTT TCAATAGGA TATGATTATG GGGCGGCAAC CTTTAAACAT	14820
CAATTCATTT TTGATATTCC GCTCGAACCA CTGCCACTAA TTTTACACTA TATTTCTCAG	14880
GATAAGCCTT GGAATCAATT TTCTGTTGGA CGTCTAAGAG AAGTTTGGTG GGAATACTCT	14940
TTGATGGATT GGTCTGTTAT TTTAAATGAA TGGTTTTCAA AGAGTGTGAA GTACCCTAGT	15000
AAATCACAAA TATTTAAGTT GCAATGTGTT AATTTAACGA ATTCTTGGTG TGTCGAGAAA	15060
ATCGATTATT TGGCGGACCA ATTGCCAGAA GTTCATTTTC ATATTGTTGC TTATACAAAT	15120
ATGGCAAATG AACTACTAGC TTTAACCGGT TTTCTAATG TTACCGTATA TCCAAATGCC	15180
TTACCAATGT TATTGGAACA AATAGTAATA GCTTCAGATT TGTATTTGGA TTTGAATCAT	15240
GATCGAAAAT TAGAAGATGC ATATGAGTTT GTGCTTAAGT AAAAAAACC AATGATAGCT	15300
TTGACAATA CTTGCTCTGA AAATCTTTCT GAGATTTTAT ATGAAGGTAT CTATCCAAGC	15360
TCCATTCGGA AAAAAATGGT TGCAGCAATC AGATCTTACA TGAGGTAGAG AACAGTATGA	15420
GAAAATCAAT AGTATTAGCG GCAGATAATG CCTATCTTAT TCCTTTAGAG ACGACTATAA	15480
AGTCTGTATT GTATCACAAT AGAGATGTTG ATTTTATAT TCTCAACAGT GATATAGCTC	15540
CTGAATGGTT TAAATTATTG GGGAGAAAAA TGGAAGTTGT GAATTCCTACA ATTGCGAGTG	15600
TACACATTGA TAAAGAACTT TTTGAAAGCT ATAAAACAGG ACCTCATATA AATTATGCTT	15660
CTTACTTTAG ATTTTGTGCG ACAGAACTGG TTGAATCTGA TAGGGTATTG TATCTGGATT	15720
CCGATATCAT TGTAACGGG GAACTAGCTA CTTTCTTTGA GATAGATCTC AAAGGATATT	15780
CAATTGGTGC TGTGATGAT GTCTATGCCT ATGAAGGACG AAAATCTGGA TTTAATACTG	15840
GTATGTTACT AATGGATGTT GCAAAGTGA AAGAACATTC TATTGTCAAT AGTTTATTGG	15900

584

AATTAGCGGC	CGAGCAGAAT	CAAGTTGTTT	ATCTTGGGGA	TCAGAGTATT	TTAAATATTT	15960
ATTTTGAGGA	TAATTGGCTA	GCCTTAGATA	AAACATATAA	TTATATGGTG	GGTATTGATA	16020
TTTATCACCT	TGCTCAAGAA	TGTGAACGTC	TAGATGACAA	TCCACCTACA	ATTGTTCACT	16080
ATGCTAGTCA	TGATAAACCT	TGGAATACAT	ATAGTATATC	TAGACTACGT	GAATTATGGT	16140
GGGTTTATAG	AGATTTGGAT	TGGTCAGAGA	TTGCTTTTCA	ACGTTCCGAT	TTAAATTATT	16200
TTGAAAGAAG	CAATCAGTCT	AAAAACAAG	TGATGCTTGT	GACATGGAGT	GCAGATATAA	16260
AACATTTAGA	GTATTTAGTA	CAACGGTTAC	CTGATTGGCA	TTTTCATTTG	GCTGCACCGT	16320
GTGATTGTTT	TGAGGAGCTG	ACCTCTCTAT	CACAGTATAC	GAATGTAACA	GTATATCAAA	16380
ATGTATTACA	TAGTAGAATT	GATTGGCTAT	TGGACGATTC	TATAGTTTAT	TTAGATATTA	16440
ATACAGGTGG	AGAGGTTTTT	AATGTAGTTA	CAAGGGCACA	AGAAAGTGGC	AAGAAAATCT	16500
TCGCTTTTGA	TATCACACGT	AAAAGTATGG	ATGATGGACT	CTATGACGGT	ATTTTCTCTG	16560
TGGAGAGACC	AGATGATTTA	GTGGATAGAA	TGAAGAATAT	AGAGATAGAG	TAATGAGTGA	16620
ATTAATTAGT	GTTGTGGTAC	CGATATACAA	TACGGGAAAA	TATTTAGTGG	AGTGTGTCGA	16680
GCATATTCTG	AAGCAAACCT	ATCAAAATAT	AGAAATTATT	TTAGTTGATG	ACCGTTCTAC	16740
GGATAATTCT	GCGGAAATTT	GTGATGCTTT	TATGATGCAA	GATAATCGTG	TGCGAGTATT	16800
GCATCAAGAA	AATAAGCGGG	GGGCAGCACA	AGCTAAAAAT	ATGGGGATTA	GTGTAGCTAA	16860
GGGAGAGTAC	ATCACGATTG	TTGATTCAGA	TGATATCGTA	AAAGAAAATA	TGATTGAAAC	16920
TCTTTATCAG	CAAGTCCAAG	AAAAGGATGC	AGATGTTGTT	ATAGGGAATT	ACTATAATTA	16980
TGACGAAAGT	GACGGGAATT	TTTATTTTTA	TGTAACAGGG	CAAGATTTTT	GCGTCCAAGA	17040
ATTAGCTATA	CAAGAAATTA	TGAACCGTCA	AGCAGGAGAT	TGGAAATTCA	ATAGCTCGGC	17100
GTTCATATTG	CCGACATTTA	ACTTGATTAA	AAAAGAATTA	TTCAATGAAG	TTCACTTTTC	17160
AAATGGTCCG	CGCTTTGATG	ATGAAGCAAC	TATGCATCGC	TTTTATCTTT	TAGCCTCTAA	17220
AATCGTCTTT	ATAAACGATA	ATCTCTATCT	GTATAGAAGA	CGTTCAGGAA	GCATCATGAG	17280
AACGGAATTT	GATCTTTCCT	GGGCAAGAGA	TATTGTTGAA	GTGTTTTCTA	AGAAAATATC	17340
GGATTGTGTC	TTGGCTGGTT	TGGATGTCTC	CGTTCTGCGT	ATTGATTTTG	TCAATCTTTT	17400
AAAAGATTAT	AAGCAAACCT	TAGAATACCA	TCAATTAACA	GATACTGAGG	AATATAAAGA	17460
TATTTGTTTC	AGATTAAAGT	TGTTTTTTGA	TGCAGAACAA	AGAAATGGTA	AAAGTTGAAA	17520
TAAAGAATTT	GTTATTTACC	ATATCACAAA	CAATGAAGGT	GAGGGGAGTG	TTTTATGACT	17580
AAGATTTATT	CGTCAATAGC	AGTAAAAAAA	GGACTATTTA	CCTCATTTCT	ACTGTTTATC	17640
TATGTATTGG	GAAGTCGTAT	TATCTCCCTT	TTGTTGACC	TAAATACTAA	AGATTTTTTA	17700

585

GGAGGTTCAA CAGCCTATCT AGCCTTCTCA GCCGCCCTAA CAGGTGGGAA TCTAAGAAGT	17750
TTATCAATTT TTTCTGTTGG ATTATCCCTT TGGATGTCCG CCATGATTTT ATGGCAGATG	17820
TTTTCTTTTT CTAAACGGTT GGGTTTAAAC TCTACGTCTA TAGAAATACA AGATCGCCCT	17880
AAAAATGTACC TGACCTTGCT AATTGCTGTG ATTCAATCCT TGGCAGTTAG CTTGAGACTG	17940
CCAGTACAAT CCTCCTATTC TGCAATATTG GTTGTCTTAA TGAATACAAT ATTGCTGATA	18000
GCAGGAACAT TTTTCTTGT TTGGTTGTCA GATTAAATG CGAGTATGGG GATTGGAGGT	18060
TCTATTGTAA TCCTCCTATC CAGTATGGTT TTAATATTC CTCAGGATGT TTTGGAAACA	18120
TTTCAGACAG TACACATTCC AACAGGGATT ATTGTGTTAC TTGCTTTATT AACCCCTGTC	18180
TTTTCTTATT TACTTGCCCT TATGTATCGA GCTCGCTATT TGGTTCCTGT TAATAAAATT	18240
GGCTTACACA ATCGATTAA ACGCTATTCT TATCTCGAAA TCATGTTGAA TCCTGCAGGT	18300
GGGATGCCCT ATATGTATGT GATGAGTTTT CTTAGTGTAC CAGCTTATTT GTTCATCTTG	18360
TTGGGATTTA TTTTCCCTAA TCATTCAGGG TTAGCGGCTT TATCAAAGGA ATTTATGGTT	18420
GGAAAGCCTT TGTGGGTCTA TGTTTATATT TCGGTCTTAT TTTTATTTAG TATCATTTTT	18480
GCTTTTGTTA CGATGAATGG AGAAGAGATT GCAGACCGTA TGAAAAATC TGGAGAATAC	18540
ATTTATGGTA TTTATCCAGG TGCGGATACT AGTCGATTTA TTAATCGATT GTCCTTCGT	18600
TTCTCAGTCA TAGGTGGTCT CTTTAATGTG ATTATGGCAG GTGGTCCCAT GCTTTTGT	18660
TTGTTTGATG AAAAGTTATT ACGATTGGCA ATGATTCCAG GCTTATTTAT GATGTTGGG	18720
GGCATGATTT TTACGATTAG AGACGAGGTC AAGGCTTTAA GGCTAAATGA GACCTATAGA	18780
CCTTTGATTT AGGAGACTTT TATGTATTAT TTTATTCCAG CTTGGTATGG GTCAGAAAGA	18840
ACATGGCATG CAGATATCAC TCCATGGTAT TTTCTCATT TTCGTCTAGA GTTTGATGAT	18900
ACCTTTCACC AGATTGGGCT CTTTCAAGAG CAAGATATAG ATTCTCGTCT ATTAGTATTA	18960
GCTTACCAGC CTCATCTACG TTATTTTTTA TATAGACATG GTGTGTTAGA AATGGATACT	19020
TATTCGGTTT TTGATGTTAT GCAAGATTTT CATAATCTCC ATACCCAAGT TTTAAGCATT	19080
AGAGATATTG AGTGGGATGA TGACTGTGAA TTTATTTATA GTCCCTTTAC GATTATCGTT	19140
CAAAAAATG GGAAGAAATT TGCTAAGGTT GAACATGGAG TTGAAGGCTT CATCAGTGAT	19200
ATACAGTATT TTGAACCAAA TGGTCAAATA CATATGCACC ATATCGTGA TGATCGTGG	19260
TTTGTATCGA GCATTATCTT TTTTGAAGAT GGGCAAGCAG CCTATCAAGA ATATCTGAAC	19320
CTCAAGGGAG AGTGGCAATT CAGAGAGCGT TAAAAAGAAG GAGGACAGGT AGAAGTCAAT	19380
CCAATTTTGG GTTATCGCTT TAAATGCTT ACCTATCAAA ATATGGGAGA TCTGGTGGCA	19440

586

GAATTTTGTG AGAATTATCT GCAAACGTAT GTGAAGGATC AGGATATTTT TATGCTTCCT	19500
TCTCATTCTC ATCATGACCA GTTGGTACTA GATCGTTTAC CTAGTACTAA TCCTAAACTG	19560
TTGAGTCTGT TCATTGGACG TAATCCTCAA GATACCTTTA GGGATTTAGA TGTAACTTTT	19620
GAAAAATCGG ATTTGATTTT GGTGGATAGA GAGGATAGTT TACGATTGTT GCAGGAGTTG	19680
TATCCTGAAC GAATGCATCA ATGTTATCAT TTATCATCTT TTGACACCCG ATTACGATTG	19740
GGACGAAGCC AAAC TAAGAA AGAATCCATC ATTTATTTTC AACTGGATT T GAGCAGGGG	19800
ATTGATAATC AAGCTCTGCT TCAAGTCTTG TCCTTTGTG CTGAAAATAA GGATACTGAG	19860
GTGATTTTGT GAGCCTTTGC TGCTAGTCAG GAGCAATGA ATGAGGTGA AGGGATTGTT	19920
GAGTCTTTCA TCCAAGAAAA CATTCAATCC GAAAACTGG GAAAGGCGAT TGATTATGGT	19980
GATGCAGAAA ATCCTCTGGA AGAAAAATA CACCAGGACT TACGCTTACA GTTTGTTAAC	20040
TTGAATGATG AGTTAGATTT GATAAAAACA CTAGAATTTG TCCGTTTGAT TGTGGATTTA	20100
AATAGACATC CTCATCTCTA CACACAGATT GCTGGGATTA GTGCAGGAAT TCCTCAAATC	20160
AACCTAGTTG AAACCGTCTA TGTGTAACAT TAAAAAATG GTTATTTGTT AGCAGATGTT	20220
ACAGAATTTT CTAAGGCTGC ACATTATTAC ACAGATAGGT TGAAGGAGTG GAATGAGTCC	20280
TTGATATATT CAATTGATAA GATTAAGGAG CACACAGGAC AACAATTTCT TGGAAAATTA	20340
GAGAAATGGA TAGAGGAGGT TAAAAATGTC AAAGGAACCT AATATTTTAC AGATAGGACT	20400
TGCCAATTGG GAAAACTACT ATGACATACC TGAAAAATG AGTTGGTATT ATTTTACCC	20460
AAACTCATCA AAAGCCCTTC GTGAAATAAT TGAAAAAGAG GATATTAACC GTTTTCATGC	20520
AGTTTTAATA GAAGATGGTC AGTATTCCAG AGACTTATTT TCCTATGTAA AATATTTTGA	20580
ACCTTATACT TTATTTTATA ACCAGAATCT ACAAATAAAT GATAGAGAGG TTGTGGATTT	20640
TCTAAAAAAA CGATGTGCAC AAGCAATTGA TTTTTTAAGT CCCCACAAC TAATCAATGA	20700
TTTAAGTAAA TCTCTTTTGT GCGGTGGGTA TGGTGATAAA CTCTTTCCCTC CGACGATACA	20760
AGTCAATCCA AATTTTACAG GAGCTATTTT TTATCAAGGA TTGGATTATG TCAGTTTGGG	20820
AGGTGAGTTT GGGCAAGATT TTGCCCAGCT TGCCTATTGG GCTTATAATA TTATGGTGCA	20880
AAAACACTC CCTATTGAGT TGTGGCTTGA ATATGAGAAG GAAGGCAATT GTGACTTTTCG	20940
TTTAGTAATC CGTAAAAATGT GGAGTGGGTC TGTGATGAT TTCTTTGAAG AAGTAATAGT	21000
ATCTGAAAAA GACTTGGAGC AAGCACTTTT TATGGATAGT CGAGACGGAG ACTACTTTCT	21060
CTCGATATCT GTTGAAGCAA GAGGTCGTGG AACTATCAAA CTAGGTAATC TTCACCAACG	21120
ATGGAGTCGA AAACAATTTG GTAAGTTTGT ACTTGGTGGG AATATCCTAC ATGATTCCAA	21180
CCGTGATGAA ATAAACTATT TCTTCCATCC AGGTGATTTT AAACCGCCTT TGACTGTCTA	21240



587

TTTTCAGGT TATCGACCTG CAGAAGGATT CGAGGGTTAC TTTATGATGA AAACCTCTGG	21300
ATGTCCCTTC ATTTTATTTT CTGATCCACG TTTAGAGGGG GGAGCTTTT ATCTCGGAAC	21360
GGATGAGCTA GAGGGAAAAG TAAAGGATAC GATCACTCAC TATCTTGATT ATTTAGGCTT	21420
TGATCATAAG GATTTGATTT TATCAGGTCT TTCTATGGGA ACGTTTCCGG CTCTCTATTA	21480
TGGTGCTTCT TTTGAACCCC ATGCCATCAT AGTTGGTAAG CCCTTGGCTA ATTTAGGAAC	21540
TATAGCTAGT CGTGGACGTT TGGACGCACC GGGTGTCTCT AACTTAGCTT TTGATTGTTT	21600
AATTCATCAT ACAGGTGGGA CAAGTTCTCA AGATATGACG GAGTTGGATC AGCGTTTTTG	21660
GAAAAATTTT AAACAAGCAA ATTTTTCAAA GACAACCTTT GGTTTATCCT ATATGAAAGA	21720
TGAAGAAATG GATCCACAAG CCTATGAACA ATTAGTGTCT TATCTGTGTA ATACAGGTGC	21780
GAAGATTTTA TCTAAAGGAA CTGCTGGACG ACACAATGAT GATACAGATA CCAATATTTT	21840
TTGGTTTTTG CACTTTTATA GAATGGTCTT AGAGACTGGT TTTGGAAGGG AGAAAAGATG	21900
ATTATTACAC AGAGACAGTC TATTCATTGG GGAGAAGTTG GTGGGACTTA TATGTATGGA	21960
ACAACTGTAT CTTATTACCC TGACAAAAGT GTTCGTCTGT ATAATCCTCT ATTGCCATCT	22020
GGTGAGATTC TAAAGACTTG GTTTTCTAGT GTCAATTACC AGGCTGCACG AACCCAACCT	22080
CAGCTTCCCT TATTA AAAAG AAAGCAGGAG TATCAACTAT CACTGGTTTT TGA CTGTCAG	22140
CCTGAAAATG GAGTTTATAC CAAGATAACT TTTTTCGACC GCTATCGTGA TATTTTAGAA	22200
AAAAAGGTAG AAAAAGTGAA AGATTTTATA TTTACTTATC CAGAAGATAG TTATACTTAT	22260
CGAGTTTCTC TTTTAAGTGC TGGATTTGAG TCCTTAACTT TTTATCATTT TTCTATCAAG	22320
GAGATCAGAA GTGTTTAGAC GTTTAGGTCA AGATTTCCAG CTTAGGAAAG TGA AAAAGAT	22380
TTTAAAGCAG ATTAATGCCC TGAAAGGCAA GATGTCCTCT CTTTCGGATC AAGAATTAGT	22440
AGCTAAAACA GTAGAGTTTC GTCAGCGTCT TTCCGAGGGA GAAAGTCTAG ACGATATTTT	22500
GGTTGAAGCT TTTGCTGTGG TCGGTGAAGC AGATAAGCGG ATTTTAGGGA TGTTCCTTA	22560
TGATGTTCAA GTCATGGGAG CTATTGTCAT GCACTATGGA AATGTTCTGT AGATGAATAC	22620
GGGGGAAGGT AAGACCTTGA CAGCTACCAT GCCTGTCTAT TTGAACGCTT TTTCAGGAGA	22680
AGGAGTGATG GTTGTGACTC CTAATGAGTA TTTATCAAAG CGTGATGCCG AGGAAATGGG	22740
TCAAGTTTAT CGTTTTCTAG GATTGACCAT TGCTGTACCA TTTACGGAAG ATCCAAAGAA	22800
GGAGATGAAA GCTGAAGAAA AGAAGCTTAT CTATGCTTCG GATATCATCT ACACAACCAA	22860
TAGTAATTTA GGTTTTGATT ATCTAAATGA TAACCTAGCC TCGAATGAAG AAGGTAAGTT	22920
TTTACGACCG TTAACTATG TGATTATTGA TGAAATTGAT GATATCTTGC TTGATAGTGC	22980

588

ACAACTCCT CTGATTATTG CGGGTTCTCC TCGTGTTTCAG TCTAATTACT ATGCGATCAT	23040
TGATACACTT GTAACAACCT TGGTCGAAGG AGAGGATTAT ATCTTTAAAG AGGAGAAAGA	23100
GGAGGTTTGG CTCACTACTA AGGGGGCCAA GTCTGCTGAG AATTTCTAG GGATTGATAA	23160
TTTATACAAG GAAGAGCATG CGTCTTTTGC TCGTCATTG GTTTATGCCA TTCGAGCTCA	23220
TAAGCTCTTT ACTAAAGATA AGGACTATAT CATTCGTGGA AATGAGATGG TACTGGTTGA	23280
TAAGGGAACA GGGCGTCTAA TGGAAATGAC TAACTTCAA GGAGGTCTCC ATCAGGCTAT	23340
TGAAGCCAAG GAACATGTCA AATTATCTCC TGAGACGCGG GCTATGGCCT CGATCACCTA	23400
TCAGAGTCTT TTTAAGATGT TTAATAAGAT ATCTGGTATG ACAGGGACAG GTAAGGTCCG	23460
GGAAAAAGAG TTTATTGAAA CTTACAATAT GTCTGTAGTA CGCATTCCAA CCAATCGTCC	23520
GAGACAACGG ATTGACTATC CAGATAATCT ATATATCACT TTACCTGAAA AAGTGTATGC	23580
ATCCTTGGAG TACATCAAGC AATACCATGC TAAGGGAAAT CCTTTACTCG TTTTGTAGG	23640
CTCAGTTGAA ATGTCTCAAC TCTATTCGTC TCTCTGTTT CGTGAAGGGA TTGCCATAA	23700
TGTCCTAAAT GCTAATAATG CGGCGCGTGA GGCTCAGATT ATCTCCGAGT CAGGTCAGAT	23760
GGGGGCTGTG ACAGTGCCTA CCTCTATGGC AGGACGTGGT ACGGATATCA AGCTTGTTAA	23820
AGGAGTCGCA GAGCTCGGG GCTTGATTGT TATTGGGACT GAGCGGATCG AAAGTCAGCG	23880
GATCGACCTA CAAATTCGTG GCCGTTCTGG TCGTCAGGGA GATCCTGGTA TGAGTAAAT	23940
TTTTGTATCC TTAGAGGATG ATGTTATCAA GAAATTTGGT CCATCTTGGG TGCATAAAAA	24000
GTACAAAGAC TATCAGGTTT AAGATATGAC TCAACCGGAA GTATTGAAAG GTCGTAATA	24060
CCGGAAACTA GTCGAAAAGG CTCAGCATGC CAGTGATAGT GCTGGACGTT CAGCACGTCC	24120
TCAGACTCTG GAGTATGCTG AAAGTATGAA TATACAACGG GATATAGTCT ATAAAGAGAG	24180
AAATCGTCTA ATAGATGGTT CTCCTGACTT AGAGGATGTT GTTGTGGATA TCAATTGAGAG	24240
ATATACAGAA GAGGTAGCGG CTGATCACTA TGCTAGTCGT GAATTATTGT TTCACTTTAT	24300
TGTGACCAAT ATTAGTTTTT ATGTTAAAGA GGTCCAGAT TATATAGATG TAACTGACAA	24360
AACTGCAGTT CGTAGCTTTA TGAAGCAGGT GATTGATAAA GAACTTTCTG AAAAGAAAGA	24420
ATTACTTAAT CAACATGACT TATATGAACA GTTTTTACGA CTTTCACTGC TTAAAGCCAT	24480
TGATGACAAC TGGGTAGAGC AGGTAGACTA TCTACAACAG CTATCCATGG CTATCGGTGG	24540
TCAATCTGCT AGTCAGAAAA ATCCAATCGT AGAGTACTAT CAAGAAGCCT ACGCGGGCTT	24600
TGAAGCTATG AAAGAACAGA TTCATGCGGA TATGGTGCCT AATCTCCTGA TGGGGCTGGT	24660
TGAGGTCACT CCAAAAGGTG AAATCGTGAC TCATTTTCCA TAAAAGGAGA AAATATGACA	24720
ATTTACAATA TAAATTTAGG AATTGGTTGG GCTAGTAGCG GTGTTGAATA CGCTCAAGCC	24780

589

TATCGTGCTG GTGTTTTTCG GAAATTAAAT CTGTCCTCTA AGTTTATCTT TACAGATATG	24840
ATTTTAGCCG ATAATATTCA GCACTTAACA GCCAATATTG GTTTTGATGA TAATCAGGTT	24900
ATCTGGCTTT ATAATCATT CACAGATATC AAAATTGCAC CTACTAGCGT GACAGTGGAT	24960
GATGTCTTGG CTTACTTTGG TGGTGAAGAA AGTCACAGAG AAAAAAATGG CAAGGTTTTA	25020
CGTGTATTCT TTTTGGACCA AGATAAGTTT GTAACCTGTT ATTTGGTTGA TGAGAACAAG	25080
GACTTGGTTC AACATGCCGA GTATGTTTTT AAGGGAAACC TGATTCCGAA GGATTACTTT	25140
TCTTATACGC GTTATTGTAG CGAGTATTTT GCTCCCAAGG ACAATGTTGC AGTCTTATAC	25200
CAACGAACTT TTTATAATGA AGACGGGACT CCAGTCTATG ATATCTTGAT GAATCAAGGG	25260
AAGGAAGAAG TTTATCATT CAAGGATAAG ATTTTCTATG GAAAGCAAGC TTTTGTGCGT	25320
GCCTTTATGA AATCTTTGAA TTTGAATAAG TCTGATTGG TCATTCTCGA TAGGGAGACA	25380
GGTATTGGAC AGGTTGTGTT TGAGGAAGCA CAGACAGCAC ATCTAGCGGT AGTTGTTTAT	25440
GCGGAGCATT ATAGTGAAAA TGCTACAAAT GAGGACTATA TCCTTTGGAA TAACTATTAT	25500
GACTATCAGT TTACCAATGC AGATAAGGTT GACTTCTTTA TCGTGTCTAC TGATAGACAA	25560
AATGAAGTTC TACAAGAGCA ATTTGCCAAA TATACTCAGC ATCAGCCAAA GATTGTTACC	25620
ATTCCTGTAG GCAGTATTGA TTCCTTGACA GATTCAAGTC AAGGGCGCAA ACCATTTTCA	25680
TTGATTACGG CTTACAGTCT TGCCAAAGAA AAGCACATTG ATTGGCTTGT GAAAGCTGTG	25740
ATTGAAGCTC ATAAGGACTT ACCGGAAC TAACCTTTGATA TCTATGGTAG TGGTGGAGAA	25800
GATTCTCTGC TTAGAGAAAT TATTGCAAAT CATCAGGCAG AGGACTATAT CCAACTCAAG	25860
GGGCATGCGG AACTTTGCGA GATTTATAGC CAGTATGAGG TCTACTTAAC GGCTTCTACC	25920
AGCCAAAGGAT TTGGTCTGAC CTTGATGGAA GCTATTGGTT CAGGTCTACC TCTAATTGGT	25980
TTTGATGTGC CTTATGGTAA TCAGACCTTT ATAGAGGATG GGCAAAATGG TTATTTGATT	26040
CCAAGTTCAT CTGACCATGT AGAAGACCAA ATCAAGCAAG CTTATGCCGC TAAGATTTGT	26100
CAATTGTATC AAGAAAATCG TTTGGAAGCT ATGCGTGCCT ATTCTTACCA AATTGCAGAA	26160
GGCTTCTTGA CCAAAGAAAT TTTAGAAAAG TGGAAGAAAA CAGTAGAGGA GGTGCTCCAT	26220
GATTGAACTT TATGATAGTT ACAGTCAAGA AAGTCGAGAT TTACATGAAA GTCTAGGCGC	26280
TACTGGTCTT TCTCAACTTG GAGTGGTCAT CGATGCAGAT GGTTTTCTGC CTGATGGTCT	26340
GCTTTCTCCT TTTACCTATT ATCTAGGTTA CGAGGATGGA AAACCTCTCT ATTTTAATCA	26400
AGTTCCCGTT TCAGATTTTT GGGAAATTTT AGGAGATAAT CAGTCTGCTT GTATTGAAGA	26460
TGTGACGCAG GAGAGGGCTG TCATTCATTA TGCTGATGGA ATGCAGGCTC GCTTGGTTAA	26520

590

ACAGGTAGAC TGGAAAGACC TAGAAGGTCG AGTACGTCAG GTTGACCACT ACAATCGCTT	26580
CGGAGCTTGT TTTGCTACAA CGACTTATAG CGCAGATAGC GAGCCGATTA TGACAGTTTA	26640
CCAAGATGTC AATGGTCAAC AAGTTTACT GGAAAACCAT GTGACGGGTG ATATCTTATT	26700
GACTTTGCCA GGTCACTCCA TGCCTTACTT TGCAAATAAA GTTGAATTTA TCACCTTCTT	26760
TTTGCAAGAT TTGGAAATAG ATACCAGTCA GCTTATCTTT AATACTCTAG CGACTCCTTT	26820
CTTGCTTCC TTCCATCATC CAGATAAATC TGGCTCGGAT GTCTTGGTAT GGCAGGAACC	26880
TCTCTATGAT GCCATTCCAG GTAATATGCA GTTGATTTTC GAAAGTGATA ATGTGCGTAC	26940
TAAGAAGATC ATCATTCCAA ATAAGGCGAC TTATGAGCGC GCTTTAGAGT TAACGACGA	27000
GAAATACCAT GATCAGTTT TGCCTTGGG TTATCATTAC CAGTTCAAAC GTGATAATTT	27060
CCTAAGACGA GATGCCTTAA TCTTGACCAA TTCAGATCAG ATTGAGCAAG TAGAAGCAAT	27120
CGCAGGAGCC TTGCCTGATG TCACTTTCCG TATTGCAGCG GTGACAGAGA TGTCTTCTAA	27180
GCTCTTAGAC ATGCTTTGCT ATCTAATGT GGCCCTTAC CAGAACGCTA GTCCACAGAA	27240
GATTCAGGAG CTGTATCAAC TGTCGGATAT TTAATTGGAT ATAAACCACA GTAATGAGTT	27300
GCTACAGGCA GTGCGTCAGG CCTTTGAGCA CAATCTCTTG ATTCTTGGCT TTAATCAGAC	27360
GGTGACAAT AGACTTTATA TCGCTCCAGA CCATCTATTT GAAAGTAGTG AAGTTGCTGC	27420
TTTGGTTGAG ACCATTAAAT TGGCCCTTTC AGATGTTGAT CAAATGCGTC AGGCACTTGG	27480
CAACAAGGC CAACATGCAA ATTATGTTGA CTGGTGAGA TATCAGGAAA CCATGCAAAC	27540
TGTTTTAGGA GGCTAACATG TCAGAGGAAG ATTTATTTTA CAAAGACGTT GAAGGCCGCA	27600
TGGAAGAGTT GAAACAAAAA CCCATCAAGA AGGAAAAAGA AACCCGAGGG GAAAAGATTA	27660
GTAAGACTTT TCACTTTTA CTGGCTTGA TGATCTGAT TGGTTTGCTC TTTACTTTGC	27720
TGGGAATTTT GAGGTAGATC TATGATTGAA ATACTAATTG TTTTAGCTAT TATCCTATCT	27780
CTTGCTTTGA TTGTATTGGT AACTATACAA CCCCGTCAAA ATCAACTATT TTCCATGGAT	27840
GCCACTAGTA ATATTGGTAA ACCAAGCTAC TGGCAGAGCA ACACCTTGGT CAAGGTGCTC	27900
ACTTTATTGG TGAGTTTGGC TTTATTTATT CTACTATTAA CCTTTATGGT GATTACTTAT	27960
AAATAAAAGA AAACCTCAGA TATTCACCTT TTGTGGATTG GTCTGAAGTT TTCTTTTTTA	28020
TACTCAATGA AATCAAAGA GCAAACCTAGG AAGCTAGCCG CAGGCTGCTC AAAACACCGT	28080
TTTGAGGTTG TAGATATAAC TGACGAAGTC AGCTCAAAAC ACCGTTTGA GGTGTAGAT	28140
ATAACTGACG AAGTCAGCTC AAAACACCGT TTTGAGGTTG TGGATAGAAC TGACGAAGTC	28200
AGCTCAAAAC ACCGTTTGA GGTGTGGAT AGAACTGACG AAGTCAGCTC AAAACACCGT	28260
TTTGAGGTTG TGGATAGAAC TGACGAAGTC AGCTCAAAAC ACCGTTTGA GGTGTGGAT	28320

591

AGAACTGACG AAGcCAGTA ACATATATAC AGCAAGGCGA CGCTGACGTG GTTTGAAGAG	28380
TATTACTGTC TATATTTTTG GTAAAAATCA ACTTTTACTT GGATGAAGGT TTTGGCTTCA	28440
CGTAGGAGTT GAAGAAGGGT GCGCGGGTTC TCAAATTCCT CTCTGTCTT GGGCAGACTG	28500
CGGTTCGGGA AGACTTCCAG ATAACGTTCA ATTTTCATCTA GCAAATCAGA AGCAGGATTG	28560
GTCTGGCTCA GTTGACCTGC AATTTTGGAA AAGAGTTGCG CTAAGATCAG GCTTTCACTG	28620
CGGGCAAGGT GACAAGTGT AATCTGTTGG GCCATGTTTC TCAGGATACG ACTTTGTGCG	28680
TGTCTCATCT CAAAGTAGTG GATATGGTAG TCTGTCTGGT GAAAGAGGTG GTCAGAGTGA	28740
TCCAAATAGA CCAGTCTGAG GGCTTCTTTC AAAAGCGTGT CTAATTCTGC TACCAGCTGT	28800
GCTCGGTTGC GTCCGTCTCC TCTGGATAAA TAGTATTTGA AGCGCTGGAG GATATCTTTT	28860
AACTTTTCTT CCACCAGCGT GTGGTAGTGC TGGATTTCTT CTTCTCGTGA AGGCATATAG	28920
AGATTAACAA GCAAGGCAAA TCCTGTACCA ATAGCAAAGA GAAGGAATTC ATTGACTAGA	28980
AGGTCTGGAG AGGTTGACTC TTGAACCAAG AGATGGCTAA CCAAACAGT GCTTGGTGTG	29040
ATGCCAATTT CCCAGCCCAT CTTGTAGGCT AAAGGAACGT AGAAGGCCAG ATAGAGGCCG	29100
AGACTCCAGA TATGAAATCC GCTCAAGTGA AAAGCTAGAA CACCGATAGC CAGAGCTAGA	29160
AGCATAGAAA AAAGACGATT GCGAGCCAGT TTTAAAGTAC TTCTACCGCT ATCAGATAGG	29220
CTCAAGAGAG CGATAATTCC AGCCGAAACT GCTGACGAAA GATTGAGAAA ATAAGCAAGC	29280
AGGCAGGCAA GACAGGTAGC TAAGATGAGC TTGGTCGTAC GTTGGCTAAT AGACATAAGA	29340
ATTTCTAAT AAGTTAGAAT AAAAGCGTAA AAGACAAGAC ATGAGCAGGC TTGCCTTGAT	29400
GAGTTATTTT TTACGGGTTG CTGCCATTTC GGCAACGGCG GTAAAGAGGA CATCTGTAGA	29460
AGAGTTAAGG GCTGTTTCAC ATGAGTCTTG GATGACACCA ATCACAAAAC CAACCCCAAC	29520
AATTTGTATG GCAATATCGT TAGAAATACC GAAAAGGCTA CAAGCAACTG GGATAAGAAG	29580
GAGGGAACCT CCGCAATAC CTGAAGCATC ACAGGATGAG ATAGCTGCTA CCACACTGAG	29640
GACAAAGGCT GTGGCAAAGT CAACAGGAAT TCCAAGAGTG TTAAGTGCAG CAAGGGTCAA	29700
AAGGTTAATG GTAATCGCTA CTCCAGCCAT ATTGATAGTA GAACCGAGTG GGATAGAAAC	29760
AGAATAGGTA TCTGGGTTGA GTCCAAGGTC ATGGCAGAGT TTCATGTTGA CAGGAATGTT	29820
AGTCGCAGAA CTACGAGTGA AAAAGGCTGT CACACCGCTG ACACGGAGGC AGTTCCAAAC	29880
TAGAGGGTAA GGATTGCGTC TCATAAAGAA GAAGGCAATC AAAGGGTTGA CCACAGGGGC	29940
AACAAAAAGC ATAGTCGTTA CTAATAGAAC CAATAAAATA CCGTAGTTGG CAAGGCTTCC	30000
GACTCCCTTG TCAGAAATGG TTTTAAAAAC AAGACCAAGG ATTCCAAATG GAGCCAGATT	30060

592

GATGATCCAT TCGACAATTT TAGAAGTCAC GTCAGCGATA GTTTTATAGCA ATTCTTGACT	30120
ATTTTACTG GCTTCTCTCA TAGCGATTCC AAAAATGACT GCGCAAGATA AGATTCTAAT	30180
ATAGTTAGCA GTAAGCAGGG CGTTGACTGG GTTGTCAACC AGTTTGAGCA AGAGGTTGCT	30240
GAGAACCTGC CCAATCCCAT CTGGTGGTGC AATTCAGTA TTGGCACTAT TTGGGGTAAT	30300
TTCAATAGGG ACGATGAAAT TTGCTAGTAC AGCTACAAGA GCAGCGGCGA AAGTCCCTAT	30360
CATAGGATAT ACAAGAAAAC AACAGTTTTC ATATTGCTAT CTTGTCCCTT TTGATGTTGG	30420
GAAAGGGCAT TGGCAACGAG AGCAAAGACT AGGATAGGAG CAACAGCTTT TAGACCTCCA	30480
ACGAATAAAT CCTCGAGTAG CCCAATCCCT GAGAGATTAG GAAGGGTCAG TCCTAGGATT	30540
CCCCACAAAG CATACCAATC AAGATACGCT TGACAAGGCT TGCCTTATTC CAAGCATGAA	30600
TGATTCTTTT CATAATAATC TCCTTTTGT GTAGTGATTA TGATTATAGT ATAAATGATA	30660
GACAAAATCA AGAATTTTCT GTCTATTTT TGAATATTTA TGGAGAATGA GACTGATGAA	30720
AATATGGTAT AATGAAATAA AGGAGTTTTA TATGCAAAAA TTTATTCAGG CTTATATTGA	30780
AAAGCTAGAT GTGACAACCA TTATCGAGAA TATCTAACC AAGGTCATTT CTCTTTTACT	30840
GCTTTTAATT GTATTTTATA TTGCTAAAAA AATGCTTCAT ACCATGGTGC AGAGAATTGT	30900
CAACCTTCT CTAAAAATGT CTCGTCATGA TGTGGACGC CAAAAACCA TCTCACGTTT	30960
ACTAGAAAAT GTGTTTAATT ATACGCTATA TTTCTTTTTA CTCTACTGCA TTTTGTGAT	31020
TTTAGGTTTG CCAGTTTCTA GTTTGCTGGC TGGAGCTGGT ATTGCTGGGG TAGCGATTGG	31080
TATGGGAGCC CAAGGCTTTC TGTCTGATGT CATCAATGGC TTTTTCATCC TCTTTGAACG	31140
TCAACTGGAT GTGGGAGATG AGGTCGTTCT GACAAATGGA CCGATTACTG TATCGGGTAA	31200
GGTTGTCACT GTGGGAATTC GTACGACACA GCTTCGTAGC GAGGAGCAAG CCCTTCACTT	31260
TGTCCTAAC CGAAATATCA CAGTTGTTAG CAATTTCTCA CCGACAGACT AGACCTGTTA	31320
TTTTAAGTAA TTTGTGGTAC AATAGAGGGA GTTTAATAAG GAGAAAAGAT GGTTTTAGAA	31380
AAGCAGTTGG GCAATCGTTG TACCTGGATA GACCTAGACC TAGGAAAGTT GAATAAACTA	31440
GAAGACCTTT CTGAAATTTA CGGTTTGGAC AAGGAAACCA TTGAATACGC ACTGGATAGA	31500
AACGAGCGCG CCCACATGGA CTACCACCGT GAAAGTGAGA CGGTTACCTT TATCTATAAT	31560
GTCTTAGACG TAAAAAAGGA CAAGGCCTAC TATGAGACTT TTCCCATGAC CTTTATTGTC	31620
GAGCATCGTC GCCTGATTAC CATTAGTAAT ACCAAGAACG CCTATGTCAT TGAACAGATG	31680
ACTCGTTATC TGGAGAACCA TGACACGCTT TCGATTTATA AGTTTCTCTT TGCCAGTCTG	31740
GAAATCATCA GCAATGCCTA CTATCCTGTC ATTGAGCAGA TGGACAAGAG TAGGGATGAG	31800
GTCAATGACC TCTTGCGCCA GCGAACTACC AAGAAAAACC TCTTTGCTCT GTCTGATTG	31860

593

GAGACTGGTA TGGTTTATCT GACGGCAGCT GCCAAACAAA ATCGGATTTT GTTAGAGCAT	31920
ATTCAAGGTC ATGCCTTGTA TCGTAGTTT GATGAGATTG AGAGAGAACA GTTTGATGAT	31980
GCCATGATTG AGGCTCATCA GCTGGTATCC ATGACAGACC TAATCTCTCA GATTTTACAG	32040
CAGCTTTCAG CCTCTTACAA CAATATTCTA AACATAATC TGAATGACAA TTTGACAACC	32100
TTGACTATCA TTTCAGTCTT GCTAGCTGTT TTGGCAGTCG TGACAGGCTT TTTCGGAATG	32160
AATGTTCCCT TACCTTTAAC AGATGAGCCC CATGCTTGGC TCTATATCAG TTTGGCTAGT	32220
GCAGGTTTGT GGATTGTTTT ATCCTTGTTA CTAAGGAAAA TTGCGAAAAA AAGTTAAGAA	32280
AAGGAGCCAG AATGGCGATT GAAAATTATA TACCAGATTG TGCTGTGGAA GCAGTCTATG	32340
ATCTGACAGT CCCAAGCCTG CAGGCGCAGG GAATAAAGGC TGTTTTGGTC GATTTGGATA	32400
ATACCTCAT TGCTTGGAAC AACCCTGATG GAACGCCAGA GATGAAGCAA TGGCTACATG	32460
ACCTTCGGGA CGCGGGTATT GGCATTATCG TAGTGTCAAA TAACACCAAA AAACGCGTTC	32520
AACGAGCAGT TGAGAAATTT GGGATTGATT ACGTTTACTG GGCCTTGAAG CCCTTCACAT	32580
TTGGTATTGA CCGTGCTATG AAGGAATTCC ACTATGACAA AAAGGAAGTG GTCATGGTTG	32640
GTGACCAACT CATGACAGAT ATACGAGCAG CCCACCGTCC AGGGATTTCG TCAATTTTAG	32700
TCAAACCCCTT GGTCCAACAT GACTCAATCA AAACGCAGAT TAACCGAACT CGTGAGCGTC	32760
GTGTTATG	32768

## (2) INFORMATION FOR SEQ ID NO: 72:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14872 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

CCAGTCACAA AGAAATTGAG CGCGTTCAGC TGAGGATGCA CTATGATGCA AGCTACATTT	60
CATTTGATGG GATATTAAGA AAGGAGATTT TCATGACACT TTTAGATGTA AAACACGTTT	120
AAAAAATTTA TAAAACACGT TTTCAGGGCA ACCAAGTAGA AGCCCTCAAG GATATTTCACT	180
TTACCGTAGA AAAGGGTGAC TACGTTGCCA TCATGGGTGA GTCTGGTTCT GGTAAATCAA	240
CTCTTCTCAA TATTCTAGCT ATGTTGGATA AACCAAGTCG TGGTCAGGTT TACTTGAATG	300
GAACTGACAC CGCAACTATT AAAAATTCAC AGGCTTCTAG TTTCCGGCGT GAAAAGCTAG	360
GATTTGTCTT CCAAGACTTT AACTTGCTAG ATACTCTGTC TGTAAAGGAC AATATCTTGC	420

594

TTCCGCTTGT CTTGTCAAGA AGACCTATAA CGGAGATGAT GAAGAAATTG GTGGTGACAG	490
CTGAGAATCT GGGTATTAAC CAATTGCAAG AGAAGTACCC TTACGAGATT TCTGGTGGTC	540
AGAAACAGCG TGTAGCAGTA GCGCGGCCA TCATCACAGA ACCTGAAATT CTCCTTGCGG	600
ACGAGCCAAC AGGAGCCCTT GATTCCAAGT CATCTGCAGC CTTACTTGAT GTCTTTAATG	660
AAATCAATGA GCGTGGGCAA ACCATCCTCA TGGTAACCCA CTCAACAGCA GCTGCTAGCA	720
GGGCCAAGCG TGTTCCTTTT ATCAAAGACG GCATTCTTTA CAACCAAATC TACCGTGGAG	780
AGAAGACAGA GCGTCAGATG TTCCAAGAAA TCTCTGATAC CTTGACTGTC ATGGCAAGCG	840
AGGTGAATTA GTATGTTTCG ATTAACCAAT AAGTTAGCGG TATCGAACTT GATTAATAAC	900
CGCAAACCTT ACTATCCCTT TGCCTGGCT GTTCTCTTGG CAGTCACCAT CACCTATCTC	960
TTTTACTCCC TAACCTTCAA TCCAAAGATT GCGGAAATCC GTGGAGGAAC CACCATTCAA	1020
GCAACACTTG GATTTGGTAT GTTTGTCGTT ACCCTTGCGT CACCATTTATC GTCCTCTATG	1080
CCAATAGTTT TGTCATGAAA AACCGTTCCA AGGAACTGGG TATATATGGC ATGTTAGGCT	1140
TGGAGAAGCG CCATCTAATC AGTATGACCT TTAAGGAGTT AGTGGTATTT GGGATTCTAA	1200
CTGTTGGAGC GGGTATCGGT ATTGGAGCCT TGTTTGACAA GTTAATTTTC GCTTTCCTGC	1260
TCAAACCTAAT GAAACTGAAG GTTGAGCTGG TTGCTACCTT CCAAATGAAT GTTGTCTATG	1320
CAGTACTTGT TGTCTTTGGA TTGATTTTCC TAGGCCTCAT GTTCTCTGAAT GCTCTTCGAA	1380
TCGCCCCGAT GAATGCCCTC CAGCTCTCGC GTGAGAAAGC AAGCGGAGAG AAAAGAGGTC	1440
GCTTCTTACC TCTCCAAACG ATTCTTGGTT CCATAAGTTT AGGGATTGGC TATTATCTTG	1500
CCCTTACGGT AACCGATCCT CTTACAGCCC TAACAACCTT CTTCCTAGCT GTTTTGCTGG	1560
TTATCTTTGG TACTTATCTA TTGTTTAATG CAGGGATTAC AGTCTTCTTA CAAATCTTAA	1620
AGAAAAACAA GAAATACTAT TACCAACCTA ATAACCTCAT ATCTGTTTCC AACTTGATTT	1680
TCCGTATGAA GAAAAATGCC GTTGGACTAG CAACCATCGC TATTTTGTCA ACAATGGTTT	1740
TGGTAACCAT GTCAGCAGCG ACAAGCATTT TCAATTCCGC AGAAAGCTTT AAAAAAGTTC	1800
TAAATCCTCA TGATTTTGGG GTTTCAGGGC AAAATGTTGA AAAAGAAGAT TTGGACAAAC	1860
TCTTGAGCCA GTTTGCAAGT GACAAAGGTT ATAGTGTCAG AGAGAAAGAA GTACTTCGTT	1920
ACAGTAACCT TGGTATTGCA AATCAAGAAG GAACCAAGTT AACTATTTTT GAAAAAGGAC	1980
AAAACCGTGT CCAACCCACA ACAGTTTTCG TGGTATTTGA CAAAAAGAT TATGAAAAATA	2040
TGACTGGTCA AAAACTGTCT CTATCAGGAA ATGAGGTCGG TCTCTTTGCC AAAAATGACG	2100
GACTGAAAGG ACAGAAAGCT CTAACCTTAA ATGATCATCA ATTTTCTGTC AAAGAAGAAT	2160
TTAATAAAGA TTTTATTGTG AACCATGTTT CAAATAAGTT TAATATCTTG ACTACTGATT	2220



595

ACAATTACCT TGTTGTTCTT GATTTACAAG CCTTTTGGGA TCAATTCCCA GATTTCGGCTA	2280
TCTATAATCA GTTTTACGGT GGTATGAATG TAAATGTCAG TGAAGAAGAA CAACTCAAGG	2340
TCGCTGAGGA GTATGAAAAC TACCTCAATC AATTTAATGC TCAATTAGAC ACAGAAGGTA	2400
GCTATGTTTA TGGTAGCAAT CTAGCAGATG CTAGTTCTCA GATGAGTGCC CTCTTTGGTG	2460
GTGTCTTCTT TATCGGTATT TTCCTATCCA TTATCTTTAT GGTTCGGAAT GTTCTGGTCA	2520
TCTACTACAA ACAAAATTCT GAAGGCTACG AAGACCGTGA ACGCTTTATT ATCTTGCAGA	2580
AAGTCGGTTT GGACCAAAAG CAAATCAAGC AAACCATCAA CAAACAGGTT TTAAGTGT	2640
TCTTCCTTCC TTGCTCTTT GCCTTCATAC ATCTCGCCTT TGCCTACCAT ATGCTTAGCC	2700
TGATTTTAAA AGTGATTGGT GTAGTGGATA CGACTATGAT GTTGATTGTG ACCTTGTCTA	2760
TCTGCGCTAT CTTCCTCATC GCCTATGTGC TGATTTTCAT GATTACTTCA AGAAGTTATC	2820
GCAAGATTGT GCAAAATGTA AAAAGATACC TCGACTTCAA AATCGAGGTA TTTCTTGTAT	2880
TCTAAATGCT GAAAAGTTGT CCGAGCAGGA AGGTAAGTCC CATCGTCAAG AGACCAATAG	2940
CAAGGTCCG AATCATAGCT GTTTTGGTTG GGGCTTTTCC AAGTCTAGCA CTTGTGTAAC	3000
CAGTGAGAAG AAGGGCCACA CCGACAATAA GGACGGTAGC AGGGATGCGG TAATCACTTG	3060
GAAAAATGGT CACTGACAGC ATTGGAGGCA AACTTCTAAG GAAAAAGGCA ACGAAGCTAG	3120
AAATGGCAGC GTGCCAAGGA TTGGTAAATT CTTCACTC AATCCCATAT TTTCTCTCTA	3180
CCAGAGCCTT GAGTGGATTT TTAAGAAAGA TCTTATTGGT CAAGAGTTGG GCAGAAGTTT	3240
TGAATCTCC ATTTTGGATA TAAGCAGCAT AGAGGGATTT TTTGGCTAGT TCCCTATCTT	3300
GGTCTAGCAA GAGTTTTTCT CGCGAAACGG CAGCTTCTC GGTATCTTTT GGAGTTGAAA	3360
CGGATACATA TTCTCCACCA GCCATTGAAA AGGCACCAGC TAAGATAGCC GTAAAACCTG	3420
ATAAAAAGAT AATCCAGATA TTGGTCGTGG CACTGGCAAC TCCGATAACC ACACCAGCAA	3480
TGGAAATAAT TCCATCGTTA GCATCAAGAA CACCCGCACG CAGGATATTT AAACGACCTG	3540
CAAAATTTGA ATCAATTTCTG TGATTTGTTT CTGACGCTAA ATTTCAAGTT CAAGTTAGCC	3600
ATCAAGAAGT CTTCTCTGGG TGAATTGTAG TCCAAGCATT TTTTAGGATA GTTGTTAATC	3660
CACTTTTTGA TGAATGCGAC TTCTTTGGGA GTCATTTTCT TGGTTCCCTT AGGTAACCAT	3720
CTACGAATGA GCCTGTTGTG ATTCTCATTA GTTCCCTTTT CCAAGAGGC ATAGGGATGT	3780
GCATAATAAA TGTGCTCCTC AGAAAATACA TTAGACAAGC GATTGAATTC CGTTCCATTA	3840
TCTGCCGTGA TGGAAAGAAT CTTGTGTTGT TTTAAGATGA GTTTTAGAGC CTGATTGACC	3900
ACATCAGCAC TTTTATTGG AATCAATCGG ATGATCTGAT GTCTACTTTT TCGATCCGTC	3960

596

AAGACAAGCA AGCAGTAGTT TTTCGCTCTC GTAAGTAGAA CTGTATCAAT CTCATAATGC	4020
CCATTCTCCA AGCGAAAATT GATAGCTTCA AGCCGCTGTT CGATGGATTG ACCAGCAGGT	4080
TTAAAGTTGG TGCTGGCCTG TTCTTAAGC GCTTTTCCTT TTCTAGGGTA AAGCAGATCC	4140
TGTTTGCTTA ACCCCAATTT TCCATGATGA ATCCAATAGT AAATGGTTGA AATTTCCACG	4200
TTAACCCCTT TAGCCATCAC CATCATTTCA GGCGAAAATT TTTGGTTATG ATAGTGGAGA	4260
ATCTTTTCCT TTAGTTCCTT GGTCAAGCTT GATTTCTTGA CCGAGCGCTT GCGATTGTTT	4320
TCATAAGACT GTTGAGCATA GTCGGCAGAA TAAACCTCTT TGAAGCGCCC TTTTCCAAGA	4380
CATTGTGCGA CTGTCCCACG CTTGATTTCA GTGTGGATAG TTTGAGGAAC TTTTCCAAGC	4440
AGAGAGGCAA TTTCTCTATT TGATTTCCCT TCTTTTTTCC ATCTTTTCGAT TAAGCGACGG	4500
CTATCGATTG TCAAATGTTT GCCTTTTGTA GTATAATGGT TTTGCATCTC TGTGCCTTTC	4560
TTGTGTTTGT GGTGAACAA CAAGTATAAC ACAGAGGTGT TTTCTTATGC CTACAAGAGC	4620
TATCGGCTAG TTGAACCATC TAATTTTATG GAGGGCTGGG TGGCTAACTT CATTATAGAA	4680
CTTTCATTTA CGAACATATA GTAAAATGAA ACAAGAACAG AACAAATCGA TCAGGACAGT	4740
AAAATCTATT TCTAACAATG TTTTAGAAGC AGAGGTGTAC TATTCTAGTT TCAATCTATT	4800
ATATTTTGT TTTTATCAA AAAATACTTT ACAAGTTCTT AAAACATGA TATAGTAATA	4860
AAGCTTAGAA AATGAGATGA TGTTTTCTAG CAAATATAAA CCCGAGTAAA AAATGCCTAC	4920
GGACAGGCAG GGTGAATGC CGAAGCGTGG TTGAAAAGCC ACATTATTGA TAGGGTTAAA	4980
AGCCTACTTT TATAAGTTGA TGTTAGGACA CTTGTCTTAA TTCATAAATT TTTAGTGTGG	5040
TGAAAGCACA CGTCATCTTG TGAAACGATC AATAAAGTAC GTAATATTTG CTAAGTAGAGA	5100
GTTAGGAAAC ATCGGGAACA GACATACTCA ACAGAAACCA AAATAAACAC GTCAGAAGAT	5160
TGCAGAGCAG GTGAAAACCT GCTCTTTTTT CATGAGTCAA CCTTTAGTTC CTTAGTTTTC	5220
ATAAGGTCCT AAAAATATTG AAAGGAGTAT GTTTTGAAAG AGTTAGATCA AAACCAAGCC	5280
CCAATTTATG AGGCCTTGGT GAAGTTACGC AAGAAAAGGA TTGTTCCCTT TGATGTTCCA	5340
GGTCACAAGC GTGGACGGGG AAATCCAGAA CTTGTGGAAC TCTTAGGAGA AAAATGTGTA	5400
GGCATTGATG TCAATTCGAT GAAACCTTTG GATAATTTAG GCCATCCTAT TTCGATTATT	5460
CGTGATGCAG AGGAGCTGGC TGCAGATGCT TTTGGAGCTA GCCATGCCTT TCTAATGATT	5520
GGTGAACAA CTTTCATCGGT GCAGACTATG ATTCTGGCAA CCTGCAAGGC AGGAGATAAG	5580
ATTATTCTGC CACGAAATGT CCATAAATCT GCTATCAATG CGTTGGTCTT ATGTGGTGCC	5640
ATTCCCATCT ATATCGAGAT GAGTGTAGAT CCTAAGATTG GTATCGCTTT AGGTCTTGAA	5700
AATGACCGAG TAGCACAGGC CATAAAGGAC CATCCAGATG CTAAGGCTAT CCTAATCAAC	5760

597

AATCCTACTT ACTACGGCAT CTGTTTCAGAC CTAAAGGGGT TGACAGAAAT GGCTCATGAA	5820
GCTGGCATGA TGGTTTTAGT AGATGAAGCC CACGGAGCGC ATTTGCATTT CACTGATAAA	5880
CTTCCAATTT CTGCTATGGA TGCAGGGGCT GATATGGCAG CAGTTTCCAT GCATAAGTCT	5940
GGTGGGAGTT TGACCCAAAG CTCCATTTTA CTTATCGGGG AGCAGATGAA TTCTGAATAC	6000
GTTTCGTCAGA TAATTAACCT GACCCAGTCT ACATCTGCCT CTTACTTGTT GATGGCTAGT	6060
TTGGATATTT CACGTCGCAA CTTGGCCCTT CGTGGTAAAG AGTCGTTTGA GAAAGTCATT	6120
GAGCTATCTG AGTATGCCCG CCGTGAATC AATGCTATCG GTGGCTACTA TGCCTACTCA	6180
AAAGAGTTAA TAGACGGTGT TTCGGTTTGC GATTTTGACG TAACTAAGCT GTCAGTTTAC	6240
ACTCAGGGTA TTGGCTTAAC AGGTATCGAG GTTTATGACC TCTTGCGAGA CGAATACGAC	6300
ATTCAGATCG AGTTTGGTGA TATCGGCAAT ATCTTGGCCT ATATTTCCAT CGGCGACCGC	6360
ATCCAAGACA TCGAGCGCTT GGTGGTGCT CTGGCTGATA TTAAGAGACT CTATTCAAGA	6420
GATGGAAAAG ATTTGATAGC AGGAGAATAT ATTCAGCCCG AGTTAGTGCT GTCTCCGCAA	6480
GAAGCCTTCT ATTCAGAAAG AAAAAGTTTA ACTTTGGATG ATTCTGTTGG ACAGGTCTGT	6540
GGAGAATTTG TTATGTGTTA CCCTCCAGGT ATTCTATCT TGGCTCCTGG TGAACGCATT	6600
ACACGAGAAA TTGTCGACTA TATCCAATTC GCCAAGGAAC GTGGTTGCTC CCTCCAAGGG	6660
ACGGAAGATC CAGAGGTCAA TCATATCAAC GTTATTAAGA GAAAGACAAA CTATAAGAAA	6720
AGTCAATAGT TTTATCTAAA CTATTTCTTA TTTCAATTTG ATGATTTGGC GATGATTTTA	6780
GAGCACGGCA AAAAGCCCTT GAATTAGAAG CGGTCAATCG CTTAATTTCT ATCAGCTTAT	6840
CAAATCCTGC CTCAAGCCTT TTCTGAGGAT TAGGGTAGCG TGTCAAGAGT TGGTAGGTAT	6900
ATTCTGAATG CTTTCCAACG ATTTTATCCA ACTCAGGAAA GATGATATCA AGACAACGAG	6960
TGTATTGTAC TTTCCAATCA GACTGTTTTT TCTTGAGACG ATGAATATGT CTAGCCAGTA	7020
TTTTTAGTTC TACTTGCCGA TTATCGTGTT GAAATTGTTT ACGATTGGGG TCAGAAAGAA	7080
GTTTAAGAGC GATGCCATGA GCGTCTTTCT TATCCGTTTT AGTTTTGCGA AGTGATAATG	7140
ATTTGGCAAA TTTCTTGATG AGCAAAGGAT TGTAGGTGTA AACTTTATAT CCTTGTTTAT	7200
GCAGGAAGTT CAGTAGATTA AAGGCATAAT GTCCGGTATT TTCAAGAGCG ATGAGACAGT	7260
CTTGGTTGAG CTGTCGAAGA GACAGATCTA AGAGTTCAAA ACCAGCTTTA TTATTTGAAA	7320
AAGTGAGTGG TTTAAGAACA GTTTTTCTTG GAACATTCAA GGCTGTAACA TCGTCTTTAT	7380
TTTTAGCGAC ATCAATGCCC ACATAAAGCA TGGGAGTATC TCCAGATATA GTATTTCAAG	7440
TCTACTGGGT TATCCACGAA CTTTTTGCCT TGTTACCTTA GACGAGATAA AACGTCTATG	7500

598

CGTTATCAAA CTCATTACCA ATTGAAACAA AAAACTGTGG TTAGAGCCTT TCGGAAATCG	7560
TCAAGCGATT GGAGGAAATG AACTAATCCA CAGTGGCTTA TTCCAAGTAT ACCACTTGGG	7620
CTTTGGCAGT AGCTAACTGC GCTAAATATA ATATAAGGAG AAATAGATGG ATTTATGGTT	7680
TTCTGAAGTT CATACTCCAG ATGTCAAATT GTCTCTGAGA ACAGCCAAGC AACTTTACGC	7740
TGGA AAAAGT GAATGGCAGG ATATCGAAGT CTTGGATACG CCAGCTTTTG GGA AAATACT	7800
GATTTTAAAT GGCCATGTCT TGTTCTCAGA TGCGGATGAT TTCGTCTACA ATGAAATGAC	7860
CGTTCACGTT CCCATGGCTG TCCACCCAAA TCCAAAGAAA GTATTGGTTA TTGGGGGTGG	7920
TGACGGCGGT GTTGCCCAAG TATTAACCCCT CTATCCTGAA CTGGAGCAAA TTGATATTGT	7980
GGAACCGGAT GAGATGTTGG TCGAGGTCTG TCGTGAGTAT TTCCCAGACT TTGCTGCAGG	8040
GCTAGATGAT CCTCGTGTTA CCATTTACTA CCAAAATGGG CTACGCTTTT TCGGAAACTG	8100
CGAAGATGAT TACGATATTA TCATCAACGA TGCGACAGAT CCATTTG GCC ATACGGAAGG	8160
ACTCTTTACC AAGGAATTCT ACGGCAATAG TTATCGAGCT CTGAAGGAAG ACGGCATCAT	8220
GATTTACCAG CATGGGAGTC CCTTCTTTGA CGAGGATGAG TCGGCCTGCC GAAGCATGCA	8280
CCGCAAGGTC AATCAAGCCT TTCCAATCAG TCGGGTTTAT CAGGCCCATATA TTCCAAGTAG	8340
CCCAGCTGGC TATTGGTTGT TTGGATTTGC ATCGAAAAAA TACCACCCTG TCAAAGATTT	8400
TGACAAGGAA GGCTGGAAAA AACGCCAGCT TTTACAGAA TACTACACTG CAAACTTACA	8460
CGTGGGAGCC TTTATGTTGC CCAAGTATGT TGAGGACATT TTAGAAGAAG AGGAAGGAAA	8520
AAAATGAGTC GTTTACTAGT TATTGGTTGT GGGGGCGTTG CCCAAGTTGC TATTTCAAAG	8580
ATTTGTCAAG ATAGCGAAAC ATTTACAGAG ATTATGATTG CTAGCCGTAC CAAGTCAAAA	8640
TGCGATGACT TGAAAGCGAA GCTAGAAGGC AAAACAAGTA CTAA AATTGA AACTGCAGCA	8700
CTTGATGCTG ACAAGGTTGA AGAAGTGATT GCCCTGATTG AAAGCTACAA ACCAGAAGCT	8760
GTTTTGAATG TAGCTCTGCC TTATCAAGAT TTAACCATTA TGGATGCTTG TTTGGCAACA	8820
GGTGTCACT ATATCGATAC AGCCAACTAC GAAGCAGAAG ACACAGAAGA CCCTGAGTGG	8880
CGTGTATCT ACGAAAAACG TTGTAAGGAA CTTGGTTTTA CAGCCTACTT TGA CTACTCA	8940
TGGCAGTGGG CTTATCAAGA GAAATTCAAA GAAGCAGGCT TGACTGCTCT TCTTGGTTCT	9000
GGTTTTGACC CAGGTGTAAC TAGTGTCTTT TCAGCTTATG CCCTCAAACA CTATTTTGAT	9060
GAAATCCATT ATATCGACAT TTTAGACTGT AATGGCGGTG ACCACGGTTA TCCATTTGCA	9120
ACCAACTTTA ATCCAGAAAT TAATCTCCGT GAGGTTTCTG CGCCAGGTTT TACTGGGAA	9180
GATGGGAAAT GGGTCGAAGT CGAAGCTATG TCTATCAAGC GTGAGTATGA TTTCCCTCAA	9240
GTTGGACAAA AAGATATGTA TCTCCTTCAC CATGAAGAAA TCGAATCATT GSCCAAGAAC	9300

599

ATTCCAGGTG TCAAACGCAT TCGTTTCTTT ATGACTTTTG GTCAATCTTA CTTGACGCAC	9360
ATGAAATGTC TTGAAAATGT TGGACTCCTT CGTACGGATA CCATTAACTT TAACGGCCAA	9420
GAAATTGTTC CAATTCAATT TTTGAAAGCC TTGCTTCCAG ATCCTGCCAG TCTTGGGCCA	9480
CGTACAGTCG GAAAAACCAA TATTGGATGT ATCTTTACAG GTGTCAAAGA CGGTGTCAAA	9540
AAGACTATCT ATATCTACAA TGTCTGCGAC CATCAGGAAT GTTACGCAGA GGTGGTTTCG	9600
CAAGCTATTT CTTATACGAC AGGAGTTCCA GCCATGATTG GGACAAAATT AGTCATGAAC	9660
GGAACCTTGA AACAAAGCTGG AGTGATAAC CTTGAGGAGT TAGATCCAGA TCCATTCATG	9720
GAAGCTTTGA ATGAGTATGG TTTGCCATGG GTTGTGGTTG AAAATCCACA AATGGTGGAC	9780
TAATGAAGTT AGAACAAGTA CCAACACCAG CCTATGTTAT TGACTTGGCC AAGTTAGAAG	9840
CTAATTGCCG CATTCTACAA TATGTACAAG AAGAGGCCGG TTGCAAGGTC TTGCTTGCCC	9900
AGAAGGCATA TTCCCTCTAC AAAACTTATC CCTTGATTAG CCAGTATCTA TCAGGTACGA	9960
CAGCTAGTGG ACTCTATGAG GCCAAATTGG CAAGGGAAGA ATTTCCCTGGT GAAGTCCATG	10020
TATTTGCCCC TGCTTTCAAG GATGCAGACT TGGAGGAATT GCTAGAGATA ATGGACCATA	10080
TAGTCTTTAA CTCAGAGAGA CAGTTGCGTA AACACGGTCC GCGTTGTCTGA GAGGCTGGTG	10140
TCAGTGTTGG TTTGCGCCTC AACCCCTCAGT GTTCAACTCA AGGcAGATCA CGCGCTCTAT	10200
GACCCTTGTG CACCAGGTTT TCGCTTTGGA GTTACTATAG ACAAGATTCC GAGTGATTTG	10260
CTAGATTTGG TTGACGGACT TCATTTTCAT ACCCTTTGCG AGCAGGGAGC AGATGATTTA	10320
CAACAACCTT TGAAAGCAGT AGAAGAACAG TTTGGTCCCT ACTTACATGA GGTAAAATGG	10380
CTCAATATGG GTGGTGGTCA TCATATTACA AGAGAAGGTT ACGATGTGGA TTTGCTGATT	10440
TCAGAAATCA AGCGTATCCG AAAAACTTAC AATCTTGAAA TCTATATCGA GCCTGGTGAA	10500
GCCATTGCGC TTAATGCCGG TTATTTAGCA ACTGAGGTAT TAGATATTGT AGAAAAAGGT	10560
ATGGAAATCT TGGTTTTAGA CGCCTCTGCG ACCTGCCATA TGCCTGATGT ACTTGAGATG	10620
CCCTATCGTC CACCTTTGAG AAATGGCTTT GAGTCACAGG AAAAAGCCCA TACCTACAGA	10680
CTTTCTTCTA ATACCTGTCT GACGGGCGAT GTGATTGGTG ATTATAGTTT TGAAAAATCCA	10740
GTCCAAATCG GAGACAGACT TTATTTTCAA GACATGGCCA TTTATTCTTT TGTCAAAAAT	10800
AATACCTTTA ATGGTATTGG ATTGCCAAGT CTCTATCTCA TGGACGAACA GGGAGACTGT	10860
AGCTTACTCA AAGCTTTTGG CTATCAAGAC TTAAAGGGA GATTATCATG ATGGACAGTC	10920
CAAAAAAATT AGGCTATCAC ATGCCAGCAG AGTACGAACC CCATCATGGT ACCCTCATGA	10980
TATGGCCGAC TCGACCAGGA TCATGGCCTT TTCAAGGAAA GGCTGCTAAA AGAGCATTTA	11040

600

CTCAGATTAT	CGAGACCATA	GCAGAAGGGG	AAAGAGTCTA	TCTTTTGGTG	GAGCAGGCCT	11100
ATCTATCTGA	AGCCCAATCC	TATCTTGGAG	ACAAGGTTGT	TTATTTAGAC	ATTCACCA	11160
ATGATGCCTG	GGCGCTGAT	ACTGGCCCAA	CCATTCTCGT	CAATGATAAA	GGTAAGAAAT	11220
TAGCCGTGGA	TTGGGCCTTC	AATGCTTGGG	GAGGCACCTA	TGATGGTCTT	TATCAAGATT	11280
ATGAAGAGGA	TGACCAAGTA	GCCAGTCGTT	TTGCTGAGGC	CTTGGAAAGG	CCTGCTCTATG	11340
ATGCTAAACC	TTTGTACTG	GAAGGAGGCG	CAATCCATAG	CGATGGTCAA	GGAACATTC	11400
TCGTAACCTGA	AAGTTGCTTG	CTTAGTCCTG	GTCGCAATCC	TAACCTGACT	AAAGAGGAGA	11460
TTGAAAACAC	ATTATTAGAA	AGTCTTGGTG	CTGAAAAAGT	TATTTGGCTT	CCTTATGGTA	11520
TTTATCAGGA	TGAAACCAAT	GAACACGTCG	ATAATGTTGC	TGCCTTTGTT	GGTCTCTCTG	11580
AGCTTGTTTT	GGCTTGGACA	GATGACGAAA	ATGATCCCCA	GTATGCCATG	TCAAAAGCAG	11640
ATCTCGAACT	CTTAGAACAG	GAAACAGATG	CAAAAGGTTG	TCACCTCACC	ATTCATAAAT	11700
TGCCATATCCC	TGCAGTTCGA	CAAGTTGTGA	CAGAAGAAGA	TTTGGCAGGC	TACATCTATG	11760
AAGAAGGAGA	AGAAAAGCGA	TACGCAGGTG	AACGACTAGC	AGCTTCCTAC	GTAAACTTTT	11820
ATATCGCCAA	CAAGGCTGTC	TTGGTTCCAC	AGTTTGAGGA	TGTAAACGAC	CAAGTGGCCT	11880
TAGATATCCT	CAGCAAGTGT	TTCCCAGACC	GTAAAGTTGT	CGGAATACCA	GCCAGAGATA	11940
TTCTCTTAGG	TGGTGGCAAT	ATCCACTGTA	TCACCCAACA	AATTCCAGAA	TAGGAGAAAA	12000
AGATGAGAAA	TGTAAGAGTT	GCAACCATTG	AGATGCAATG	CGCTAAGGAT	GTGGCAACAA	12060
ATATCCAAAC	CGCAGAGCGT	TTAGTACGTC	AGGCTGCTGA	GCAAGGAGCC	CAAATTATTC	12120
TCTTGCCCGA	GTTGTTTGAA	CATCCCTATT	TCTGTCAGGA	ACGTCAGTAT	GACTIONTACC	12180
AGTATGCCCA	ATCTGTAGCG	GAAAATACTG	CCATTTCAGCA	TTTTAAGGTG	ATTGCTAAGG	12240
AACTACAAGT	TGTTTACCA	ATCAGTTTCT	ATGAAAAAGA	TGGTAATGTC	TTGTATAACT	12300
CTATTGCCGT	CATTGATGCA	GATGGGGAAG	TGCTGGGCGT	TTATCGAAAG	ACCCATATAC	12360
CAGATGACCA	TTATTATCAA	GAAAAATTCT	ATTTACGCCC	TGGTAACACT	GGTTTCAAGG	12420
TCTGGAATAC	TGGCTATGCT	AAGATTGGTA	TCGGTATCTG	TTGGGATCAA	TGGTTCCCTG	12480
AAACAGCGCG	CTGTCTTGCA	TTGAATGGTG	CTGAATTGCT	CTTTTATCCT	ACAGCTATCG	12540
GTTTCAGAGCC	AATTTTGGAT	ACAGATAGTT	GTGGTCACTG	GCAACGTAAT	ATGCAAGGGC	12600
ACGCAGCAGC	GAATATTGTT	CCAGTCATCG	CAGCCAAATCG	TTATGGTTTA	GAGGAGGTTA	12660
CTCCTAGTGA	GGAAAATGGC	GGACAGAGCT	CCAGTCTTGA	CTTCTACGGT	TCCTCTTTTA	12720
TGACGGATGA	AACAGGAGCT	ATTCTAGAAC	GAGCTGAAAG	ACAAGAAGAA	GCTGTTCTGT	12780
TAGCTACTTA	TGACCTAGAC	AAGGGAGCAA	GTGAACGCCT	AACTGGGGC	TTGTTTCGAG	12840

601

ATAGAAGACC AGAAATGTAT AGACAAATTA CAGATTAGTG TGGGAGAAAT GAGAGATTCA	12900
TTCTGCTAGA CTAACCTCTT ATTAGTAACT ATAAGATACT ATGGCATCTA GTAAATCGAT	12960
TTTTATGATT CGCTATTCTT GTCTATTGAT TAGTCCGTAT TTTAAATAT TAGCAAAAAA	13020
GCAAATAGCA GTAACCTCTG TCTATTGCT TTTCTTTTTT ATAGAATATA TTTCTCAATA	13080
GCACGCGCAA CGCCGTCTTC TTCGTTGCTT GAGGTAACGG CATCCGCAAG AGATTGATA	13140
TAATCGCTGG CATTTCCCAT TGCAATCCCA AGCCCTGCAA ACTGGAGCAT TTCGATATCG	13200
TTATTAGCAT CGCCCATGGC CATAATCTCT GAGGAATCAA TCTTCAAAAT CTCAGCTACT	13260
CGTGAAAGAG CAGTAGCCTT TGTCGTCCA AGCGGCATTG CTTCAATAAT GACAGGCTGC	13320
GAACGAACTC CACTGAATCG TTGGCAAAGC TCTTCAGCAA AACGCTGCTC AAAATCGTCT	13380
GTTTGTTCTT TTGTTCTTAA ACACATACCT TGAACATCC GGAACCTTCC ACTACTCGCT	13440
TCTTCAAGAG AAATTTCACT CAGGTCTGAA AATACTAGTT TAGCATCATT TTCAATAACT	13500
TGATTGGGCT TGTACCCGAG AACAAAATAA TGTGACTCGT CAAAAAGTGT CAACTGAACA	13560
TCACTCTTTT CAGCAAGGTC ATAGAGGTAT TCGATGTCAG CTGGACTCAG TTCTTTCCAG	13620
TCAACTAGAC TCCAATCACT GGTCTGGTGA GTTGAACAAC CGTTCTTAAC AATAATATAT	13680
TCGTTCTGGA GGTCAAGCTC CAGTTTTTTG TAGTAGGGGA GGACACCGAA AAGGGGGCGA	13740
CCCGTACAGA GAACCAGTTT GACACCTTTT TCAATGGCTT TGTGAATAGC AGTAATGTGT	13800
GCTTGTGGGA TTTCCTTGGC TTCATTGAGG AGGGTGCCGT CCATATCCAA GGCTAGTAGT	13860
TTAATCATAG GTCTTCTCTT TTATCTTTCG TATTATTATA GCATATTTTG GAGAAGAAAT	13920
TGATAGAAAG CTTGAGACTA ATTGATTTTA TAGTTTAAGA TGTTTTGATG ACAATTCATG	13980
ATTTGAAGAG GATATTTTCG AAAGATATGC TATACTATGT TTGTCAATGT TGCAACTAGA	14040
CAAATTAATA AACCAACTTA ATATAATAGT TTTTTGTAA GTAGGTATGA GTAGCAGATT	14100
ACTCAACTAA TCTGAAGAAT AATGGAGGAA ATATATCATG ATTTTAATGA CAAAAATAT	14160
AAATCTAACA AATGAAGAAT TAGAGCTGAT ACAAGGTGGA GCAGATCCAT ATGGTAAAAA	14220
TCCTAATGGT AGGTACGATT GGGAAATAGA ACCAGTATTA ACTCTGCTGG TTCATGGATT	14280
TTGTCCCAGA GGCACCTATG ATTCAGGATA TATTGGAGGA GGTAAATCATC TTGCAAAGG	14340
AAGTGCTGCG AGATTTTAAAG TAAAATTTAT TAGGAATATG AAGAAACAAG GGGAGAAAAC	14400
AGAGGATTTA ATATGAAAAA ACGAGCTATT CAAATTTTAC TAGCATTGTC CTTAATTTTT	14460
TACAAATCAA CTTGGTTTTG GAGGCTTTTC AATTATCTCG CAAAGCCCTA TCTACCAGCA	14520
AGTCGTGAAT TTTTTCAGAT TCTGCTTTTG ATGGAGAGCG GAGTTCTTTT CTTAGCGGTC	14580

602

ATCTATCTAC TGGTTTTTGC AGGAAAGAAA ATTTTTCATT TCAAGTGGCA GCTGAGGTAC	14640
TTCATCTACC TTTTACTGGG CTACATCATT TCATATATGT CTGACTTCCT CTTTTCGTAT	14700
TTCATATCCC TGTCTTCAAA TCAGATTCTT TTGAATGAAA CGGTAGAAAT GATGGGAGA	14760
CAGGAGTTCC CTTATGTCTT GCTCATCGTT TGCTTCATCG CCCCTATTGC TGAGGAATTG	14820
ATTTATCGAG GtGTGCTTAT GACAACCTGT TGCAAAAAC CACCTTGTA CG	14872

(2) INFORMATION FOR SEQ ID NO: 73:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10223 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

CGTGCTATCG GTCTCAAAAC CAATCTGGTC GCTATGGTCA AATCCAGTTG GAAAATCCAT	60
TCTTCTTGA GCCATCTGCT GGATTGCCAT CATCTCACC ACTCTTGTA TGCAGACCTT	120
TATCGGCATT TTCTAATACT CTTGAAAAT CTCTTCAAAAC CACGTCAACG TCGCCTTGCC	180
GTAGGTATAT GTTACTGACT TCGTCAGTTC TATCTGCAAC CTCAAAACGG TGTGAGCT	240
GACTTCGTCA GTTCTATCTG CAACCTCAAA ACGGTGTTTT GAGCTGACTT CGTCAGTCGT	300
ATCTACAACC TCAAAACAGT GTTTTGAGCT GACTTCGTCA GTTCTATCTG CAACCTCAAA	360
ACAGTGTTTT GAGCAGCCCG TGGCTAGTTT CCTAGTTTGC TCTTTGATTT TCATTGAGTA	420
TAACACAAAA GGTAGCCCAT CAGCTACCTT TTTCTATGC TTCCTCAATC AAGCGAGTAT	480
GTTCTCTCTT GATACAGCGA TTCATCAGCA TATCATCACA TCCACCATCA CGCAAAATCT	540
CTTTCGCTTC TAAACTTTCA AGTCCTAGCT GTGCCAAAAA AATCTTGGA TCAGCTTTGA	600
GAAAATCAGC CGCCACATCG GGCAGAAATT CACTGCGACG ATAAACATTG ACAATATCTA	660
CAGGAAAAGG AATTTAGCG AGGCTAGCAT AAGCCTTTTC ACCCAAGATT TCGCCACCTG	720
CCGCCTTGGG ATTGACTGGG ATGATTTTAT AGCCCCGAGC CTGCATTTCC TTTGTTACTC	780
GATTGCTGGT TGTTCCTCA CGGTCAGACA AAGCCACCAC AGCAAGGGTT TTAAGCTTG	840
CGAGATACTG ACGAATCAGC CCATCACTTG GATTGATAAA TTCTTGACTC ATAGAAATCC	900
TCCTTTTCA TCAGTATAGC ACATTTTGAA AAGGTTTGCA GAATTATACT ACAAAAAAGG	960
AGGACTAGCC CCCTTTTAT TTAGCCTCGT ACCAGGTTGC CCCTTCATTC TCATCTGCGA	1020
TAAGAGGAAC ACTGAGTTGA ATGGCTTCTT CCATGGTTTG TTTCACCAAT TTTTTCATCT	1080
CTACCAATTC AGATTTAGGC ACTTCAAGGA CGATTTCATC GTGCACTTGT AACAGCATCT	1140



603

TAGTCTGATA ACCACCTGCA ACCAAGGCTT TATCCAGCTG AATCATGGCA ATCTTGAGAA	1200
TATCTGCTGC CGAACCCTGG ATAGGTGAGT TGATAGCAGT TCGCTCCGCA AAACCACGAA	1260
TATTGAAGTT GCGCGAATTG ATATCTGGCA ACTCACGGCG ACGCTTAAAG AGGCTCTCTA	1320
CATAGCCCTT ATCACGCGCC TCCCGCACCA CTTTCATCCAT GTAGTTTTTA ATACCTGGAA	1380
AACGTTCAAA GTAGGTATCA ATGTAGGCTT TGGCTTCCTT ACGACTAATT CCCAAATTAT	1440
<hr/>	
TAGACAAGCC AAAGTCTGAA ATCCCATAAA CCACTCCAAA GTTAAC TGCCATTGC	1500
GACGGTCGTT TGCAGTCACA TCATCAGGAC GCTCAATGCC AAAGACCCGC ATGGCTGTCTG	1560
AAGTATGGAT ATCTGCCCCC TCTTGAAGG CCTTAATCAA GTGCTCATCC TTAGAAATAT	1620
GCGCCAAAAC GCGCAATTCA ATCTGTGAAT AGTCAGAGCT GAGTAGCACA CTATCCTCCC	1680
ACTCTGGCAC AAAAGCCTTC CGAATCAAGC GCCCCGTTC CAATCGGGCA GGAATATTTT	1740
GCAAGTTTGG ATCCACACTA GACAAACGCC CGGTCTGGGT CAAATCCTGC ACATAGCGAG	1800
TATGAATCTT TCCATCAGCC AAAATCCAGT CCTGCAAGCC AATTACATAA GTAGATTGAA	1860
TCTTAGCAAT TTGACGGTAA TCCAGGATTT TCTTAACAAT CGGAGCAATA GGAGCGAGAC	1920
GCTCTAAAAC ATCCACTGCT GTCGAATAAC CTGTCTTGGT TTTCTTAGTG TATTCTAGAG	1980
GAAGTCCCAA TTTCTCAAAG AGAAGCACGC CCAACTGCTT AGGCGAGTTG ACATTAAACT	2040
CCTCACCAGC CAGCTCGTAA ATCTCTTGAG TCAGTTTTTC AATGACAAGC TCATTTTCAG	2100
CCTGCATCTC AAGCAAGGTC TCTTTCTTGA CCATAATCCC AGCAATTTCC ATCTTGCGAA	2160
GGACAAAAGC CAGAGGTGTC TCCATATCAT AAAGAAGCTC TAATTGCCCA TTTTCGCTGA	2220
GTTTTTCAAG TAAAATAGGC TCTGTTTCTA CCAAACAGC AAGTTTACAA GCTAAGTGTT	2280
CCAAGAATTT CTCACGTTCA GGAATGGCCT TTTTAACACC CTTACCGTAG AAAGTTTCAT	2340
CATCAACCAA GTAAGTCTGA CCATAAAGAC TAGCGATGGT CGCAATTTCA TTGTCCTCCA	2400
CAGTCGAAAG GAGGTATTTA GCCAAACGGA TGTCAAAAGC AGGCGCCTGC AAATCCACAC	2460
CAAAACGTTG CAAAAGAACT TTAACCTTCT TAAAGTCATA AACTCTCAGA GATGTTTTTT	2520
CTAAGAAATC CTTGAAAATC GGGTCTTGCA ACAGCTCAAG CTTGTCTGTG GCATAGAGCT	2580
TATCCCCACA AGACCAGACA AATCCAACCA AATTATCCGT ATGGTAATTC TCACCAAAAA	2640
GCTCAAAGTG GAAGATAGAC TCTTCACTCA GCATATCTTG ACTGATTTGG TCAACAATAG	2700
TAAAATCCAA ACTCTCAGAC ACATCAGCTG ACGACACATT TAAAGCCTGC TTTAGCTGTT	2760
TGAAGCCCAT CTCATCGTAG AATTTCCCAA GATTTTCAAC ATCTGGACCA CTATAGACCA	2820
AGTCCTCTAA ACCAATCGCA ATCGGTGCCT TGGTATCAAT GGTGCTAGT GTTTTAGACA	2880

604

AAAAGGCCTG TTCCTTGTC	TTGATGAGAT TTTCTTCAT	CTTAGAAGTC TTCATTCCAT	2940
CAATATTTTC ATAAATCCCC	TCAAGCGAAC CATGCTCCAG	CAAGAGCTTA ATACCCGTCT	3000
TTTCACCGAC TTTGGTCACC	CCAGGGATAT TATCCGACTT	ATCACCCATG AGCGCCTTGA	3060
GATCGATAAA CTGAGCTGGT	GTGAGGCCCA TTTCTTCAT	GAGGTAATCT GGCCTAAAGG	3120
CCTCAAATC AGCCACACCT	TTCTTGAAA TTTCAACCAC	CGTATGCTCA TCCGTCAGCT	3180
GAATCAAATC CTTGTCCCCA	CTGACAATAG TAATATCAAA	ACCATCCTGC TCTGCTAGCT	3240
TATCCAGCGT CCCAATGATG	TCATCCGCCT CATACTGAGC	CAGATCATAG TGACGAATCC	3300
CCATATGATC CAGCAACTCA	CGAATGAAAG GAAATTGCTC	ACGAAACTCA TCAGGAGTCT	3360
TGGCCCGACC ACCCTTATAG	TCCGCATACA TCTCTGTCCG	GAAGGTCTGC TTTCCCGCAT	3420
CAAAAGCCAC CAAAATATGA	CTCGGCTCAA CCCGCTCCAA	TAAATGACTC AACATCAACT	3480
GAAAACCATA AATCGCATTG	GTATGCAAAC CAGCCACATT	CTTAAAACGG TCCAAC TGCT	3540
GATACAGCGC AAAAAACGCC	CGAAAAGCTA CAGAAGACCC	ATCAATCAAT AATAATTTT	3600
TCTTATCCAT ACACCCATTA	TAAAGGAAAG AATCAAAAAA	TACCATTGGG AAGAGCTAGA	3660
GCAAGTATTT TTCAAATTT	TTCCGAATAA ATAGATAGAG	CCAGAGAATT TAGTAAACCT	3720
AGATTTAAAA ATGTGCTATA	ATATAGTATA TTGAATCTAT	AATAGTACAC CTTGACTGCT	3780
AAAATATTTT TATAAATTAA	TTTGACTTTT CTGATAGAGT	TATTCACATC TTATTTCAAC	3840
TCACTATAGA AGGAGGAATA	GGAGGATTCT CAGACATCCG	GGCATCAGCC CAACTAATGA	3900
TTTGATTGCT AAGAAAATAT	TCAGCAATCC AGAAATCACT	TGTCAATTTA TTCGCGATAT	3960
GCTGGACTTG CCAGCAAAAA	ATGTGACCAT TTTGGAGGGA	AGCGATATTC ACGTATTACT	4020
CTCCATGCCT TACTCGGTGC	AGGATTTTTA TACCAGTATA	GACGTCTTGG CGGAGTTGGA	4080
TAACGGTACT CAAGTAATTA	TTGAGATTCA AGTCCATCAT	CAGAATTTTT TCATCAATCA	4140
CTTGTGGGCT TACCTGTGCA	GTCAGGTAA TCAAAATCTT	GAAAAAATTC GTCAGCGAGA	4200
AGGTGATACT CACTAGAGCT	ACAAACACAT CGCTCCTGTT	TACGCCATTG CTATCGTGGA	4260
TAGTAATTAT TTCTCAGATG	ACCTGGCTTT TCATAGCTTT	AGTATGCCCG AAGACACAAC	4320
AGGTGAGGTA TTGGCGATTA	CCAACAATGG ACAGGAAAAC	CATCTGGTTA AGATGGCATT	4380
CTTGGAATTA AAAAATACAG	AGAAACCAGC AAAGACAAGG	TTGCAAGCC ATGCTTGAG	4440
TTTTTCGGCA ACAAGCCCTT	TACCCAGCAA CCGCAACGAG	CCATTACCCA AGCAAATCAA	4500
CTGCTGGACT ACAAGAGCTG	GTCCGAGGAG GACAGGAAAA	TGTTTAGTCA ACTACATATG	4560
CGAGAAGAAC AAGTCTTGTT	AGCACAGGAC TATGCCTTGG	AAACTGCTAG GGCTGAAGGC	4620
CTTGAACAAG GACTAGAGCG	TGGGAAAGTT GAAGGAAGGG	CAGAAAGGAA ACTTTTTGCC	4680

605

TTCTAGACA TAGTACGCCA AGGTCTTCTG ACTTCTGAGG TTGCCAGCCA GCAATTAGGT	4740
ATGTCAGTAT CTGAATTTGA GGCAGTGTG TAAATGGCT CCATAATATC CATAGTGGGT	4800
AAATCCCCTA TGGATATTAT GGAGCCTATT TTGTGTAGAA AAAAAGTCCC ATATGACCTA	4860
TAATGAAAAG CGACAAAACA ACTCATTAGA AAGAATCATA TGAACAATT ACATTTTATC	4920
ACAAAATTAC TAGACATTAA AGACCCTAAT GTCCAGATTT TAAACATCAT CAATAAGGAT	4980
ACACACAAGG AAATCATCGC CAAACTGGAC TACGACGCCC CATCTTGCCC TGAGTGCGGA	5040
AACCAATTGA AGAAATATGA CTTTCAAAA CTTTCTAAAA TTCCTTATCT TGAAACGACT	5100
GGTATGCCTA CAAGAATTCT CTTAGAAAG CGTCGATTCA AGTGCTATCA CTGTTCAAAA	5160
ATGATGGTCG CTGAACTTC TGATGACGTA CAGTCATATT TCTTCTCTTT TTATTATATC	5220
ACAGTTTAA ATCTAGCTTT ACTAGATTCA CCGCTACTAT CTATTTATTC GGAAAAAGA	5280
CGAAAAAACC TGAGAATCAT CTCAGGCTTG GTCATTAAAT TTTTCTCTCA ATATCGAAAA	5340
GTGGAGAAAG TGGTCGTTT TCATGAATAC GTACGATAGC ATCCCCTAGG AGATGAGCGA	5400
TTGAAATCTG CTCAATCTTA TCAATCAAAC GCTCTTCTGG CAGATAGATG GTATCCAAAA	5460
CAACCAATTT CTTAATAGCT GATTTTGGG TATTGTCCGT AGCAGGACCA GAAAGAACTG	5520
GGTGCGTACA GCTTGATAG ACTTCAACAG CACCAGCTTC CGCAAGAGCA TCTGCCGCAT	5580
GACAAATCGT TCCAGCGGTA TCAATCATAT CATCAATCAA GATACAAGTC TTGCCTTCAA	5640
CCTTACCGAT GATATTCATA ACTTCACTAG TATTCATCTT ATCAACGCTA CGACGTTTAT	5700
CAATAATAGC GATAGATGTT TTCAAAAATT CTGCCAAGTT ACGAGCACCA GTCACCCCTC	5760
CATGGTCCGG GCTGACAACC ACATAGTCAG AACCAACCAT ACCACGACGC TCAAAATAAT	5820
CTGCAATCAG AGGAGCACCC ATCAAATGAT CCACAGGAAT ATCAAAGAAT CCTTGAATTT	5880
GCGCAGCATG CAAGTCGATG GTCAATAAAC GATCCACTCC AGCTACTTCA AGCATATTTG	5940
CGACAAGTTT TGAAGTGATT GGCTCACGCG CTCTCGCCTT TCTATCCTGA CGTGATACCC	6000
CATAGTAAGG CATGACAACA TTGACAGATT CTGCACTCGC ACGCTTCAAA GCATCTACCA	6060
TAATCAAAAT TTCAAGCAGA TTGTCATTTA CAGGCGAACT AGTTGATTGT AAGATAAAGA	6120
CGTGTTCCTC ACGGATTGAT TCTTCAATGT TGACCTGAAT CTCTCCATCT GAAAATTGGC	6180
GAACACTTGA TTTCCCAAC TCTATCCAA TCTCTGCGC CACACGTTCT GCCAATCTT	6240
TATTAGAAGA AAGGGCAAAC AGCTTTAAAT CAGAAAAAGA CATGATTTCC TCCGTATAT	6300
ATGTATAACT TGTGCTTTTC ACAAGATTTT CCATCTACCA TTGTAGCGCT TTTTGCCTA	6360
TTTTTCAATC AAAAATAAAA GAAGGGCACC ATATTGTGAC CCTTGCATCA TTCTTTTGAA	6420

606

AAATATTCTA GGTCAATCAAC TCATTGTGTT TCTCAACAAA GCAATAAGCA TGATAAAAAC	6480
CATAGAGAGC AATAGCCGTA ACCACTGGAA TCGCTAAAGG CAACTCTGTT TCCAACCTCA	6540
CAAAAGGAGA GTTAAACAAG AAGTGAGTTC CCAAGGCTAA ACCTAGAAAA ATAAGGCCCT	6600
GTTTCTTGCC AACCTTCTGT CCTTTATAGG CTCTGTAAAG CAAGTAAACA CCTACTACAG	6660
CTAGACCTGA AAAAGTCCAG TGAGAGGCAA TTCCTGAGAT GATACGCTCT AAAATTGCGG	6720
AAATAGTAAA GTCAAAGCCC TCTGGCAAAT CCGTACGAAT ATAACCAATA TCCTTAATCA	6780
TTTGAATCC CAAACCGGAA GCAATTCCAA GTAAAAACAA AGATTTTAAT TTTGCGACAG	6840
GAATCAAAGC CAAAACAAAA ACAAGTGACA ATAATTTCAA GGGTTCTTCT ACCAAAGGAG	6900
CCGCAATAGC ACTTTCAAAG GCATTTAAAA ATGGACTATC TGGGAAAAGA ACCCCCAGTA	6960
AATCATGGAT ATAAGTATTA GCAAACTAG ACAACCAGCC TGAAAGGAAC ATCCCTCCCA	7020
ATAAAGACAG AATCAAAACC TTCTTTGGCA ATTCCTCATTT TTTCCAATAC GGAAGAGAAA	7080
ATAAAGAGCC GGAATCATGT AAAAGAGAGC TAGAAAGATA GAAACTCCCA TTAGTCCATA	7140
TTCCGCACCT GACCTCGAAC CGTCCGTATA GTAGATGGTT TCATACTGTA AACCAATACA	7200
TAGCAATAAA ATAAAAATAA ATAAATATT GCTTTTCTTC ATACACTTTC TTTCTAAATG	7260
AAGTATTTAT AATTCTACGA CTGTCATACT TCCTGTATCA ACATTGTAAA TGGCACCAGA	7320
GATAATGACA TCCTCTGGTA TTAGGGGAGA CTCGATAAGC AGTTGCATAT CCTCGCGTAC	7380
ACTCTCTTCT ATATCTTGA AGGGCAAGAA GTCCTGGTCT GACACATCGA CACCCAATTC	7440
TTCTTTCAAA TACTCCTGAA AAGGTTTATT TTCAAAGGTC TGAGCACCAC AGTCTGTATG	7500
ATGCAATACC ACAATTTCTC TTGTCCCCAT TTGTTGCTCG GAAATAACTA GAGAACGAAT	7560
CATATCCTCA GTCACCTGAC CACCTGCATT CCGCAAAATA TGAGCATCCC CAAGTGCCAA	7620
ACCTAGAGCT TGCCTAACGT GCAAACGTGA GTCCATACAG GTCACAATGG CTACTCTGGT	7680
TTTAGGTTTA AGTGGCAGAT TTAAGTGCCC ATGTAGGGCA ACATAAGCCT GATTGGCTTG	7740
CATAAACTGT TCAAAATACG ACACGATTCC CTCCTTGAAA ATTTGATAGT CAAATATTTT	7800
TCCTATCTTA TCATTTTAA GAGAATTTGT CACGGATTAT GCAAAGACCT TTTTCAAGAC	7860
TTCTGAATC GTTGTCACGC CAATGACCTG AATTTCTTGA GGCAGAGTGA TTCCTGTCAA	7920
GGAATTTCTTA GGTACATAAA TCTTAGTAAA GCCCAGTTTA GCAGCTTCGT TGATGCGTTG	7980
CTCAATACGA TTCACGCGCC GAATCTCTCC TGTCAGCCCC AGTTCTCCGA CAAAACATTC	8040
CTGAGGATTA GTTGGCTTGT CTTGTAGCT CGAAGCAATA GCAACTGCAA CAGCCAAGTC	8100
AATCGCAGGT TCATCCAATT TAACACCACC AGCAGATTTG AGATAGGCAT CCTGATTTTG	8160
CAAGAGAAGC CCTGCCCGTT TTTCCAAAAC AGCCATAATC AAGCTAGCAC GGTAAAAATC	8220

607

AAGTCCTGTC GTAGTAGGCT TGGCATTTC	AAACATGGTC GGTGTTACCA AAGCCTGAAC	8280
CTCCGCCAAA ATCGGACGCG TCCCTTCCAT	GGTTACAACG ATGGAGGAAC CAGTCGCCCC	8340
ATCCAAACGC TCTTCTAGGA AAACCTGACT	CGGATTGAGT ACCTCAACCA AGCCGCCCGA	8400
CTGCATCTCA AAAATCCCAA TCTCATTAGT	GGAACCAAAA CGATTTTGA CCGCTCTCAA	8460
AATACGAAAG GTGTGGTGAC GCTCCCTTC	AAAGTAAAGC ACCGTATCCA CCATATGCTC	8520
CAACATACGA GGCCAGCCA AGGTTCTTC	TTGGTCACA TGACCTACGA TAAAGATGGC	8580
AATGTTATTG GTCTGGCCA ACTGCATGAG	TTCAGCGGTC ACTTCACGCA CCTGAGAAAC	8640
AGACCCCTGC ACCCTGAAA TCTCAGGAGA	CATGATGGTC TGGATGGAAT CAATAATGAG	8700
AAAGTCTGGC TGGATACGCT CCACTTCTGC	ACGAACACTC TGCATATTGG TCTCTGCATA	8760
GAGATAAAAC TCACTATCAA TATCACCTAA	GCGCTCTGCA CGTAGTTTAA TCTGCTGGGC	8820
AGACTCCTCC CCACTGACAT AGAGAACTGT	CCCCACTTGG GACAACTGGG TTGAGACTTG	8880
TAGGAGAAGA GTTGATTTC CAATCCCAGG	ATCCCCACCG ATAAGGACGA GACTTCTGG	8940
TACCACTCCG CCTCCAAGCA CACGGTTGAA	TTCTCCATC TCCGTCTTGG TTCGATTGAC	9000
ATTGATGGAA GTCACCTCAG CTAGTTTCAT	GGGCTTGGTT TTCTCACCTG TCAAGGACAC	9060
ACGCGCATTC TTAACCTCGG CAACCTCAAC	CTCTCCACA AAAGAAGACC AAGACCCACA	9120
GTGCGGGCAA CGTCCCAGAT ATTTAGGGGA	ATTATACCCA CAATTTTGAC ATACAAATGT	9180
CGCTTTTTTC TTTGCGATGA CAAACCTCTT	TCTATATCTC TAACTCACAC TCAATCACTT	9240
GGCAAAAATC AATCTTCTCA TTTGGCACAA	ACTGGCGCAT GAGCATTGCA TGAGCAACAA	9300
CTACCACAGT CTGATGTTCT CGATACTTAG	ACATACATTC TAGAAACCGA GACTTCATTT	9360
CCGTAGCTGT CTCATATTGA ATAGGACTAT	TAGGAAGCAA CTCCCCCTTG TTTCTAAAA	9420
ACAGTCTTCT AGCTGTTTCA AAGTTTCTA	TTCTGTTTT ATAGACCTGC CATTGATGTA	9480
ATAAAGGCTC TACTCTTAAA GGAAGACCCG	TAGCACAGAC CACATACGAA GCCGTTTCTA	9540
AAGCTCTTGT GACTGCAGAA GATACGATTA	TTTCAGCTGA CGAGAGTAAA GGATTTTTC	9600
TCAATTTCTG GACTTGCTGC CGTCCCATCT	CAGACAAGGG TGCCAAATCT ATCCCAAATC	9660
CTATATAAGA ACGCTCCTCT AACTCACGGT	AATCTGGCTC CCCATGACGT ACAAAGATAA	9720
TCTTCATTCT AGTGCCCTGT CGATCCAAAT	CCACCAGTTC GAACGCCATC AGCTGCATCT	9780
CCATCTGCAA TTAAGAAAGT AGCAAAAACA	GCCTGGACAA TACGCTCCCC AACTTCAAGA	9840
ACAACCTCTT GGTCTGTGAT ATTCTTCATC	TGCGCAAAAA TATGCCCTTC ATTTCCAGGA	9900
TTTCATAAT AATCCCCATC AATGACTCCA	ACTGAGTTAA TTAACCAACAA GCCCTTCTTA	9960

608

CGAGGATTTG AAGAACGATC ATAGAGGTAG AGAACCTCAG TCGGCTGCAT ATAAGCCTTA	10020
ACCCCTGTCG GAACCAAGAC AATCTCTCCT GGCGCAACAA CTGTACGCAC AGCAACCTTT	10080
AAGTCGTAAC CAGTCGCATG CGCTGTCTCA CGCTTGGGCA ATAAATTTTC ATCTGTAAAA	10140
CTCGAAACCA ATTCAAAACC ACGAATTTTC ATAATTTTCT CTTTTCTATT ATCATTATT	10200
CTAGATTATT CTATACTTAT TTA	10223

## (2) INFORMATION FOR SEQ ID NO: 74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16535 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

TGGTTCTGTC CTTATCGGCG CCTGTCTTG CTGCCATCG CTACACCAAC TATCTCATCC	60
GACGAAAGTA CACCAACCAC TAACGAACCC AACAAACAGAA ATACAACCAC CCTTGCCCAA	120
CCTCTTACTG ATACAGCAGC TGGCTCTGGT AAGAACGAAA GTGATAATTTC TTCACCTGGA	180
AATGCAAACG CTTCCCTAGA GAAAACAGAA GAAAAACCTG CTGCAAGCCC AGCCGATCCA	240
GCACCACAAA CTGGACAAGA TCGTTCAAGT GAGCCAACTA CTTCTACTAG TCCAGTAACA	300
ACTGAAACTA AGGCAGAAGA GCCCATCGAA GATAACTACT TCCGTATCCA TGTCAAAAAA	360
CTTCCTGAAG AAAACAAGGA TGCTCAAGGA CTATGGACTT GGGACGATGT TGAAAAACCA	420
TCTGAAAACT GGCCAAACGG AGCTTTGTCC TTCAAGGATG CCAAGAAAGA TGA CTACGGC	480
TATTACCTAG ATGTCAAATT AAAGGGAGAA CAAGCCAAGA AAATTAGCTT CCTCATCAAC	540
AATACAGCTG GAAAAAATCT AACC GGCGAT AAATCTGTAG AAAA ACTAGT TCCAAAAATG	600
AACGAAGCTT GGT TAGACCA AGATTACAAG GTTTTCTCTT ACGAGCCACA GCCTGCAGGA	660
ACTGTTGCGG TCAACTACTA CCGCACAGAT GGCAACTATG ACAAGAAATC TCTCTGGTAC	720
TGGGGAGATG TGAAAAATCC AAGTAGCGCT CAATGGCCTG ACGGAACAGA CTTTACGGCT	780
ACAGGCAAAT ATGGCCGCTA TATCGACATT CCTCTTAATG AAGCCGCAAG AGAATTTGGA	840
TTTTTATTAC TAGATGAGAG CAAACAAGGA GACGACGTGA AAATCCGTAA AGAAAAATTAT	900
AAGTTCACAG ATTTGAAAAA TCATAGCCAA ATTTTCCTAA AAGACGATGA TGAATCGATT	960
TACACAAATC CATACTATGT CCATGATATC CGTATGACAG GAGCCCAACA CGTAGGCACT	1020
TCTAGCATTG AAAGTAGCTT TTCAACACTT GTCGGTGCTA AAAAAGAAGA TATCCTCAAA	1080
CACCTCAACA TCAC TAATCA CCTAGGAAAC AAGGTA ACTA TTACCGATGT TGCAATCGAT	1140

609

GAAGCTGGTA AGAAAGTGAC CTACAGCGGA GATTCTCTG ACACAAAACA TCCTTATACT	1200
GTAGCTACA ATTCCGACCA ATTCCTACC AAAACAAGCT GGCGCCTGAA AGATGAGACA	1260
TACAGCTATG ATGGCAAAC GGGAGCTGAC CTAAGAAG AAGGAAAACA AGTTGATTG	1320
ACCCCTTGGT CACCAAGTGC TGATAAGGTT TCTGTTGTTG TCTACGACAA GAATGACCCT	1380
GACAAAGTAG TTGGAAGTGT CGCTCTTGAA AAAGGGGAAA GAGGAACTTG GAAACAACT	1440
CTAGACAGCA CAAACAACT CGGAATCACA GATTCTACTG GCTACTATTA TCAATACCA	1500
ATCGAGCGTC AAGGTAAAC TGTTCTTGCA CTCGATCCTT ACGCTAAATC TCTTGCTGCT	1560
TGGAATAGCG ACGATTCCAA GATTGACGAT GCCCATAAAG TGGCTAAAGC CGCCTTTGTA	1620
GATCCAGCTA AACTCGGACC TCAAGACTTG ACTTATGGTA AGATTCACAA TTTCAAGACT	1680
CGTGAAGACG CCGTTATCTA CGAAGCTCAT GTGCGTGATT TCACCTCAGA TCCTGCCATT	1740
GCAAAAGACT TGACCAAAC ATTTGGGACT TTTGAAGCCT TCATTGAAA ACTAGACTAT	1800
CTCAAAGACT TGGGTGTAAC CCATATCCAG CTCCTCCAG TCTTGCTTA CTACTTTGTC	1860
AATGAATTGA AAAACCATGA ACGCTTGCT GACTACGCTT CAAGCAACAG CAACTACAAC	1920
TGGGGATATG ACCCTCAAAA CTACTTCTCC TTGACTGGTA TGTACTCAAG CGATCCTAAG	1980
AATCCAGAAA AACGAATCGC AGAATTTAAA AACCTCATCA ACGAAATCCA CAAACGTGGT	2040
ATGGGAGCTA TCCTAGATGT CGTTTATAAC CACACAGCCA AAGTCGATCT CTTTGAAGAT	2100
TTGGAACCA ACTACTACCA CTTTATGGAT GCCGATGGCA CACCTCGAAC TAGCTTTGGT	2160
GGTGGACGCT TGGGGACAAC CCACCATATG ACCAAACGGC TCCTAATTGA CTCTATCAA	2220
TACCTAGTTG ATACCTACAA AGTGGATGGC TTCCGTTTCG ATATGATGGG AGACCATGAC	2280
GCCGCTTCTA TCGAAGAAGC TTACAAGGCT GCACGCGCCC TCAATCCAAA CCTCATCATG	2340
CTTGGTGAAG GTTGGAGAAC CTATGCCGGT GATGAAAACA TGCCTACTAA AGCTGCTGAC	2400
CAAGATTGGA TGAACATAC CGATACTGTC GCTGTCTTT CAGATGACAT CCGTAACAAC	2460
CTCAAATCTG GTTATCCAAA CGAAGGTCAA CTTGCCTTTA TCACAGGTGG CAAGCGTGAT	2520
GTCAACACCA TCTTTAAAA TCTCATTGCT CAACCAACTA ACTTTGAAGC TGACAGCCCT	2580
GGAGATGTCA TCCAATACAT CGCAGCCCAT GATAACTTGA CCCTCTTTGA CATCATTGCC	2640
CAGTCTATCA AAAAAGACCC AAGCAAGGCT GAGAACTATG CTGAAATCCA CCGTCGTTTA	2700
CGACTTGGA ATCTCATGGT CTTGACAGCT CAAGGAACTC CATTTATCCA CTCCGTCAG	2760
GAATATGGAC GACTAAACA ATTCCTGAC CCAGCTACA AGACTCCAGT AGCAGAGGAT	2820
AAGTTCCAA ACAAATCTCA CTTGTTGCGT GATAAGGACG GCAACCCATT TGACTATCCT	2880

510

TACTTCATCC ATGACTCTTA CGATTCTAGT GATGCAGTCA ACAAGTTTGA CTGGACTAAG	2940
GCTACAGATG GTAAAGCTTA TCCTGAAAAT GTCAAGAGCC GTGACTATAT GAAAGGTTTG	3000
ATTGCCCTTC GTCAATCTAC AGATGCCTTC CGACTTAAGA GTCTTCAAGA TATCAAAGAC	3060
CGTGTCCACC TCATCACTGT CCCAGGCCAA AATGGTGTGG AAAAAGAGGA TGTAGTGATT	3120
GGCTACCAAA TCACTGCTCC AAACGGCGAT ATCTACGCAG TCTTTGTCAA TCGGATGAA	3180
AAAGCTCGCG AATTTAATTT GGGAACTGCC TTTGCACATC TAAGAAATGC GGAAGTTTGT	3240
GCAGATGAAA ACCAAGCAGG ACCAGTCGGA ATTGCCAACC CGAAAGGACT TGAATGGACT	3300
GAAAAAGGCT TGAAATTGAA TGCCCTTACA GCTACTGTTC TTCGAGTCTC TCAAAATGGA	3360
ACTAGCCATG AGTCAACTGC AGAAGAGAAA CCAGACTCAA CCCCTTCCAA GCCTGAACAT	3420
CAAAATGAAG CTCTCACCC TGCACATCAA GACCCAGCTC CAGAAGCTAG ACCTGATTCT	3480
ACTAAACCAG ATGCCAAAGT AGCTGATGCG GAAAATAAAC CTAGCCAAGC TACAGCTGAT	3540
TCACAAGCTG AACCAACCAG ACAAGAAGCA CAAGCATCAT CTGTAAAAGA AGCGGTTCTGA	3600
AACGAATCGG TAGAAAACTC TAGCAAGGAA AATATACCTG CAACCCAGC TAAACAAGCT	3660
GAACTTCCAA ATACAGGAAT CAAAAACGAA AACAACTCC TATTTGCAGG AATCAGCCTC	3720
CTTGCGCTCC TTGGTCTCGG TTTCTTACTA AAAAATAAAA AAGAGAACTA AACTAGCCCT	3780
CCTATAGAAA AATCCCCCAA GCATTATAGC TCGGGGGATT AATTTTGTG CAATATTTGT	3840
TGTCCTAATA AACTTGATTA GGATTTTITA TTAAGCCTCT TTCATAGCAA AATAAGCTCG	3900
TACTTTGGGT GCAACTTGTG TTCCGAAGAG TTCAATAGCT CTCAGAACCCT GGTATGAGG	3960
CATAGAACCA AGCGGTAGAT GAAGCATGAA GCGGTCCAAT CCTAAATCCT CTATCATGCG	4020
AATCAATTTT TCGGCCACCT GATCTGGATT GCCAACAAAC ATGGCGCCAT TTGGCCCTAC	4080
CTGCTCCAAA TATTGCTCAT AACGCAATTC CTGCCAGTGC GGACGGTCTT TGGAAATAGC	4140
ATCCACCACT TGCTTAGTCG GATGGAATA ATCTTTCACC GCCTGCTCAC CATCTTCCGC	4200
AATCCACCCC CAAGAATGGG CTCCCACCTT CAAGTCTTTG TCAGCATGGC CCTTCGCTT	4260
CCAATCTCAC GATAAGCCTG AATCAACTTT TTAATAAAC GTGGATTACC ACCAATAATA	4320
GCATATACAA TCGGTAGACC AGCCTGAGCA ATCTTCACTG TTGATTCGAC ATGACCACCT	4380
GTAGCTATCC ACAAGGGCAA TTTGTCCTGA ACTGGACGAG GATAAACTTC TTTACCAGCA	4440
ATCGTTTGAG TCAATCGACC TTGCCAGTCT AACTTGGTCT TTTCAATTGAC TAACTGAAGC	4500
AAGTCTAATT TCTCATCAA AAGAGAGTCG TAGTCTTTCA AGTCATAACC AAACAGAGGG	4560
AAAGATTCCG TGAAAGAGCC CCTTCCAGCC ATAATCTCCG ATCGTCCATT TGACAAAGCA	4620
TCGATAGTGG CATACTGTTG GAACAAACGA ATCGGGTCCA TGCTTGACAG AATGCTGACT	4680



611

GCACTGGTCA AACGGATTTT CTGGGTATTG ACTGCCCCAG CGGCCAGAAC AATCTCTGGG	4740
GCTGATACTG CAAAATCCGC CCGATGGTGC TCACCAATCC CATATACATC CAAACCAACC	4800
TTGTCAGCCA GCTCAATCTC TGCCACCAAC TGGCGAATGC GTTCAGCATG ACTGTAAGTT	4860
TGTCCAGTCC CTTCAAGCTC CGTTATTTC CCAAATGTTG AAATTCCCAA TTCTACCATT	4920
GTGATTCTCC TTATCTATCT CTGTACTTCA ATTTGAAAAA TTATTCTAAC ACGAATCTTG	4980
AGTACAAGCA ACCGATTTGC TCATTAGAAA AAGCCTAGAT AACTAGACTT TTTTAGCTTA	5040
TTCTACCGTT ACTGACTTGG CAAGGTTACG TGGTTTGTCC ACATCGAGGC CACGGTGGAG	5100
GGTTGCAAAG TAAGCGACTA ATTGCGTTGG TACGACCATT GAAATTGGTG AGAGGTATGG	5160
ATGTACGGTC GTAAGGACGA TATCGTCGGT ATCTTTGGCT ACATTCTCTT CTGCGATAGT	5220
GAGGACTTTG GCACCACGGG CTGCGACCTC TTGGATATTT CCACGAGTAT GATTGGCAAG	5280
AACTGGATCT GACAAGAGAG CCAAAACAGG CGTTCCTTCT TCAATCAAGG CAATGGTTCC	5340
GTGCTTGAGT TCTCCTGCAG CAAAGCCTTC ACACTGGATA TAAGAAATCT CTTTGAGTTT	5400
GAGACTTGCT TCCATGGCTA CGTAGTAATC TTGACCAGT CCGATGTAAA AGGCGTTACG	5460
AGTTGTTTCA AGAAGTTCAC GAACCTTGAC TTCAATGGTT TCTTTCTCTG AAAGAGTTGA	5520
TTGATAGAC TGAGCTACGA TTGACAATTC ATGAACCAGG TCAAAGGCTT GCGCTTTAGC	5580
ATTACCATTT GCTTCTCCGA CTGCTTTTGC AAGGAAGGCA AGGGCTGCGA TTTGCGCTGT	5640
ATAGGCTTTA GTTGATGCCA CGGCAATTC AGGACCTGCG TGAAGGAGCA TGGTATAGTT	5700
GGCTTCACGT GAGAGGGTTG AACCTGGAAC GTTTGTCACT GTTAAGCTTG GAATCCCAT	5760
TTCAATTAGC TTGACCAAAA CTTGACGACT ATCCGCTGTT TCACCAGATT GGCTGATAAA	5820
GATGAAGAGT GGTTCCTTGC TGAGAAGTGG CATACCGTAG CCCCCTCAG ATGAGATTCC	5880
AAGTTCAACT GGTGTATCTG TCAATTCTTC CAACATTTTC TTAGAAGCAA ATCCTGCATG	5940
GTAAGATGTT CCAGCTGCAA GGATGTAGAT GCGGTCTGCG TCTTGAACAG CCTTAATGAT	6000
ATCTGGGTCT ACGACAACCT GACCAGCCTC ATCTGTGTAG GCTTGGATGA GTTTCCGCAT	6060
AACAGTTGGT TGCTCGTCAA TTTCCTTGAG CATGTAGTAA GGGTAAGTTC CCTTACCGAT	6120
ATCTGACAAG TCAAGTTCAG CAGTGTAGCT AGCAGGCTCA CGACGATTC CATCATAGTC	6180
TTGAACCTCC ACACTATCAG CCTTGACGAT TACCAACTCT TGGTCATGGA TTTCCATGTA	6240
TTGGTTAGTT TCACGAATCA TAGCCATGGC GTCTGACCAG ACCATGTTAT AGCCTTCTCC	6300
AAGACCAATC AAAAGTGGTG ATTTATTTTT AGCTACGTAG ATGACTTCAG GATCTGTGA	6360
GTCAACCAAG GCAAAGGCAT AAGAACCAGC GATGATGTGA AGGGCTTTTT TGAAGGCTTC	6420

612

AAGAACTGAG AGCCCTTCTT CTTCCGCAAA TTTTCCAATC AAATGAACGG CTATTTTCAGT	6480
ATCTGTCTGC CCCTTGAAGT GGTGACCTGC AAGGTATTCT TCCTTGATTT CAAGATAGTT	6540
CTCAATCACC CCATTATGCA CCAAGACAAA ACGTTCCGTC TCAGAGCGGT GTGGGTGAGC	6600
ATTGTCCTCA GTTGGTTTTT CGTGAGTAGC CCAACGAGTA TGTCCGATAC CAGTTGTTCC	6660
CTCAACACCA GCTGTCTTGG CAGACAATTC TGCAATACGA CCAACCGCCT TCACCAAATG	6720
GTTATCAGCA CCATCTAGGA CAAAAATTCC CGCAGAATCA TAGCCACGGT ATTCAAGCTT	6780
TTCAAGCCCT TGAATCAAAA TATCAGTTGC ATTTGTGTTT CCAACAACAC CAACAATTCC	6840
ACACATAGTA TATACGACAC AGGCAAGCTG TGCTTTCTCC TTAATAATTGG TATAGTCTAA	6900
TTTATCTTTT ATAGAATCAG CAAAAACAGT ATATACTTGT TTCTTTCACT TGTCAAGAGT	6960
AAAAATTGGT ATAGTTCAAA TTAAGCTCCT GTAAGCATAA AAACCTCTGAC CGATTGGGAT	7020
AATCAGTCAG AGTCCTTTTT AAAATCCATT ATTATCGCTT AATTCTTTGA ACCAGTGGCC	7080
TGATTCTTTC AGACGACGTT CTTGCGTTTC CAAGTCTAAT TCGACCAAAC CATAGCGATT	7140
TTTATAGCTG TTGAGCCATG ACCAGCAGTC AATAAAGGTC CAAATCAAGT AGCCCTTACA	7200
GTTGGCACCA TCTTCAATGG CACGGTGAAG TTCACGAAGA TGACCTTTTA CAAAGTCAAT	7260
ACGGTAATCA TCTTGAATCA TTCCATCTTG ACGGAATTTT TCTTCCCCTT CAACACCCAT	7320
ACCATTCTCA GTCAACATCC ACTCAATATT GCCATAATTT TCCTTGATAT TTTGGGCGAT	7380
GTCATAAATC CCTTGCTCAT AAATCTCCCA ACCACGGTGA GAATTGATTT TACGTCCAGG	7440
CATCACATAA GGCTCGTAAA AATGTTCTGG TAAGAGTGGG CTCTCTGGAT GCTTAGCAAA	7500
TCGAGGAGCC ATAACACGCA AAGGTGATA GTAGTTCACA CCAAGGAAGT CCACCGTATT	7560
ATCACGAATG AGTTCCAAC TTTCTCTGT AGCATCAGGT AAAAGACCGT GTTCATGCAA	7620
GATTTCTACC AACTCCTGTG GATAAGTCCC CAAGACAGAT GGATCTAAGA AAGATTGGGC	7680
CTGAAAAAGG GCCGCAATAC GAGCTGCCTT GACATCAGCA GGATGCTGGC TACGTGGATA	7740
AGCCGGTGTC AAGTTGAGGA CAATCCCAAT CTTGGAATCA GGCAAAAGTT CATGGCAAGC	7800
CTTAACAGCC CGGCTGCTGG CCAATTGTGT ATGATAGGCT ACCTTAACAG CTGCCTCTGC	7860
ATCCACCTTA TGTGGATAAT GGGCATCATA AAAATAACCA AATTCTACAG GAACGATGGG	7920
CTCGTTAAAG GTAATCCATT GATCCACTAA ATCTCCATAA GTCTCAAAAC AAAAACGAGC	7980
ATAGTCTTCA TAGGCTGAGA CTGTCGCCTT ATTTTCCCAA CCATCACCAT CCTCTTGAAG	8040
GGCAAAAGGT AAATCAAAAT GATAGAGATT GACTAACAGA CGAATTCCTT TAGCCTTAAT	8100
AGCCTCAAAG ACCTTACGAT AAAAATCCAC ACCTTGAGTG TTGACTTTTC CACAGCCTTG	8160
TGGAAAAATC CGTGACCACT GAATAGAAGT CCGAAAGGCT GTGTGACCAG TCTCTAACAA	8220

613

AAGCTCAATA TCCCGCTCCC AATTTTCATA AAAAGTCGAT GTCTTATCTG AACCAATCCC	8280
ATTATAGTAA CGATTTGGCT CCACTTGGA CCAGTAATCC CAGAGATTGT CTCCCTTACC	8340
GTCACCAGCT ACACGTCCTT CTGTCTGCGG TCCAGAAGTA GAGGATCCCC AGACAAAATC	8400
CTTTGGAAAT CTTAGCATAC ATTTACCTCT TTATCTACTC ATTTCTCCCA TTATACAGAA	8460
AAAACAAGGT AAAAAGTACT TACATTTTTT CTTTGTTTTT CTTCTGATTA TAGTTTTTAT	8520
TTCTTGCTTA GGATTTCAAG CGTTTCAAGC ACGTTATCTG CATGAACCTC AATGGTGTCA	8580
CCAGTTGCCT TGATCTTAAC TTCTACAATG CCATCGGCCG CTTTTTTTACC AACAGTGATA	8640
CGGATTGGAA GACCAATCAA GTCACATCG CTAAATTTAA CACCGACACG TTCGTTACGG	8700
TCATCTGTCA AGACTTCATA ACCAGCTCCC ATCAAGCTTG CTTCAAGTTT TTCTGTCAAG	8760
GCTTGCGCTT CTTTCATCCTT GACATTGACA GTAATCAAAT GCACATCAAA TGGTGCCAAT	8820
TCTTTAGGGA AATTGATTCC CCAAGCGTAA CGGTATTAC CTTTTGGCGT TTTGTTAACA	8880
AAGAGGCGAG CGTGTGTCTC CATCACTGCT GAAAGAAGAC GGCTGACACC GATACCGTAA	8940
CATCCCATGA TGATTGGCAC AGCAGCACCA TTTTCATCCA AGACATCTGC TCCCATGCTT	9000
GCTGAATAGC GAGTTCCGAG TTTGAAAATA TGACCGATCT CAATACCACG CGCAAAGTTA	9060
AGGACACCTT GTCCATCTGG GGAAATTTCA CCCTCACGAA CTTACCGAT ATCCACATAT	9120
TCTGCAGTAA AATCACGGCC TGGGTTCA CAAGTCAAGT GGTAGTCATC TTCGTTAGCA	9180
CCGACAACCTG CATTGCGAAC ATCTGTGACC TTACGATCTG CAATAATTTT AATATTCTCT	9240
GGCAAACCAA CTGGTCCAAG TGAACCAAAT CCTGCTTGAA CAACATTCGC CACTTCTTCT	9300
TCGCTAGCAA CGTCAAAGAA ATCTGCTCCC AAGTGATTTT TCAACTTGAC TTCGTTGAGT	9360
TGGTCAATTC CAACTAGAAG GGCTGCAACA AGCTCACCAT CTGCAATGTA GAAGAGGGTT	9420
TTAATCGTTT GTTCTTCTGG AACATTGAGG AAGGCTGCAA CTTTCATCAAT TGATTTAACA	9480
TCTGGCGTTG CAACACGAGT AACTTCTTCT TCAGCGACAA CACGGTTGCT TGCTTTGTAC	9540
TCGTTTGTG CCATTTCTAA GTTAGCTGCA TAGCTAGACT CACTTGAGTA AGCAATGTA	9600
TCTTCACCAG AGACTATCCA TTTGAGCAAT TCTGCCTTGA TTTCTTCTTG CACTTCTGCA	9660
GGAATTTTCGT CAAATGAGGC AACTGACTTG TCCAAGACAA CCCAGCGGTC AAGGTCTGTA	9720
CGAGCAGATG TAATGGCCAT AAATTCTTGG CTATCCTTAC CACCCATGGC TCCACCGTCA	9780
CCAATAATAG CCTTGAAGTC TAAACCACTA CGAGTGAAAA TACGCTCATA GGCTGCTTTG	9840
TACTCATCAT AAACACTATC CAAACTATCA TAGTTAGCGT GGAAACTATA AGCATCCTTC	9900
ATGATAAACT CACGTGTACG AAGAAGTCCA TTACGCGGGC GTTTTTCATC ACGATACTTG	9960

614

GGCTGAATTT	GATAAAGGTT	GAGTGGCAAT	TGCTTGTAAG	ATTTAACAGA	ATCACGGACA	10020
ATAGCTGTAA	AGGTTTCTTC	GTGAGTTGGA	CCTAAGATAA	AGTCTGATTT	TTACACGGTTT	10080
TTTAGTTTGT	AAAGGTCTTC	ACCATAGGTT	TCGTAACGAC	CTGATTACAG	CCACAATTCT	10140
GCACTAAGAA	GGGCTGGAGC	CAACATCTCA	ACAGCACCAA	TCTTTTCGAA	TTCTTGGCGC	10200
ATGATGTTTT	TAGCTTTTTC	AATCACACGG	TTGGCAAGTG	GTAATAAGA	ATAAACACCT	10260
GCTGAAACTT	GGCGAACATA	ACCAGCACGC	AACATAAGAG	CATGGCTGAT	AACTTGAGCA	10320
TCGCTTGGCA	TTTCGCGAAG	CGTTGGGATA	GGCATTTTAC	TTTGTTTCAT	AATATTCCTC	10380
GATTATCTAA	AAAAGAGTCG	CATAATGTCA	TTCCAAGTCA	CAGCAATCAT	CAAGACAACC	10440
ATGATGACCA	CTCCGGCCAA	GGTGACATAG	GTTCCAATTT	CTTGTTTCAA	TGGTTTGCCG	10500
CGGATGGCTT	CTAGGATATT	GAGCACAACT	TTACCACCAT	CCAAGGCTGG	AATCGGAATA	10560
AGATTAAAAA	TCCCAATATT	GATGGAAATC	ATTGCCAAGA	AGTACAAGAT	ATTTTCAATT	10620
CCATTTTtag	CAGCATCACT	ACTTGCCTTA	AAGATAGCAA	CAGGTCCACC	CAACTTGTTT	10680
AAATCTGGTT	GGAAAATCAG	ATTTTTCAGA	GCTGAGAGAA	TTCCGAGAGC	TGAGTCAGCA	10740
GCAGTTGTAA	AACCACCTAC	AAACATGGAT	AGAAAATCTG	ACTTAACCCC	CGGTTGAACA	10800
CCTAGAAGGT	AACGACCTTG	ACTATCTTTG	GGTGTAAACAG	TGACTTGTTT	GTCACCTCCC	10860
TTTTTCAGAA	TAGTCACATC	CAAAGTCGGT	GCCGCTTTAT	CTTTGGTTTC	TGTTTCCACA	10920
GCTTGGATCA	AGCTTTCCCA	GTTGCTAACC	TCATGTGAGC	CAATCTTGGT	AATTTGTGCC	10980
ATTTCTGGTA	CTCCTACCTT	GGCCAAGGCA	CCTTGGGGCA	TGATATGGAA	CTGATTGGTA	11040
TCAACATCTC	TGACACCACC	CTGCATAAAG	ATTAAAACCC	AAAAACAAC	GACACCTAAG	11100
ATAAAATTGT	TCATAGGACC	TGCAAAATTG	GTAATCAGTT	TGCCCCAGAT	AGTCGCATTT	11160
TGATATTGAA	CATCTAAAGG	TGCAATCCGA	ACCTCAGTAC	CATCTGCTTC	CACAACCGTT	11220
GCATCGTGAT	CCACTGCAAA	TGTTTTTTCT	TCTTCCAGAA	CCAATCCTTT	GATAAAGAGC	11280
TTGTCTTCAA	AATCAAACCTG	GGTCACCTGC	ATAGGGAGGG	CTGTTTGATC	CAATTTTAA	11340
CCTGAGAGAT	TGATGCGTTT	AACCTTACCA	TCATCAGCAA	GTGTCAAAC	AACAGGCGTT	11400
CCTGTCTTGA	TTTCAGTTGT	ATCATCACCC	CAACCGGCCA	TGCGGACATA	GCCACCCAGA	11460
GGCAAGATTC	GAATGGTATA	GGCCGTTCCT	TCCTTGCCAA	TGTGAGCAAA	AATTTTAGGT	11520
CCCATACCGA	TGGCAAATTC	ACGTACTAAA	ATCCCTGATT	TCTTGGCAAA	GTAGAAGTGA	11580
CCGAACCTCG	GCACCACTAC	AATAATCCCG	AAAACCAGAA	TAAAGGTTAA	AATTCGGAGC	11640
ATAGCGTTTC	CTCCGTCTTT	TGATTAAAAG	AGTCCAAATA	AGTGCATGAT	TGGAAATACA	11700
AGCAACATAC	TATCGAAACG	ATCCAAAACA	CCACCATGTC	CAGGGATAAA	TTTCCAGAA	11760

615

TCCTTAACAC CAAAATGACG TTTGATCGAA CTTTCTAGTA AATCACCAAA TTGTCCAGCA	11820
ATGCTAAAGA AAATAGCAAA GACTGACATC TTGTAAATTC CATATGGAAG AGCAACTGTA	11880
CTGTCAACTA TCATAAGGAT AATGGTTACT AAAATTGCTC CTAAAATACC ACCCAAGGCA	11940
CCCTCAAGGG TTTTATTAGG CGATACCCCTT GGTGCTAACT TTCGTTTCCC ATAGTTCATC	12000
CCAACAAGAT AGGCACCACT GTCTGTCGCC CAGACGATAC ACAAGGCTAA GAGAGCCTTG	12060
-----	
TCCAAACCTG CAACACGAGC ATCTAGTAAA GCATTAAATC CAAAGCCCAC GTAGAAGCTC	12120
ATAGCAAGAG GGAAAACCGC ATCCTCAATC GTATAAGACT TGCTAAAAAC GGTCGTTCTT	12180
AACATGATTG AAATCAAAAC ACTATAGGCA ACCACATTCC CATCAACTGG CAAAAAAGTC	12240
AGGTAATTCT CCAAGGGAAT GGTCAATGCA AAGGTTGCAA AGAGGGTCAA GAGGCCCTCC	12300
ATCGTCATGG TCTCTAGACC TCTCATCTTC AAAAGTTCAT GCATGGCTAG CATGGCTATG	12360
ATTCGATTG CTATCTGAAG CAAGAGGCC CCAATCATTAA AAATTGGTAG GAAAAATAGCC	12420
AGGGCAATCC CTGCAAAACAA GGTCTTTTTC TGTAAATCCT GGGTCATATT TCCTCCTAAA	12480
CTCTCCAAA TCGGCGATGA CGACGATTAT AGGCAAGAAT AGCTTCCTGC AAGGCCGCTT	12540
CGTCAAAATC AGGCCATAAG GTGTCCGTAA AATAAAGCTC ACTATAGGCT CCCTGCCATG	12600
GAAGGAAATT GCTCAACGT AATTCTCCAC TAGTACGGAT AATCAAGTCT GGGTCTCGTA	12660
AGTCCTTAGG CAAATGCTGA GTAAAGAGAT AGTTACCAAT CAATTCCTCT GTGATGTCAC	12720
CTGGGTTGAT TTTGGCATCT AAAACATCCT GGGAAATCAA CTTAAGCGCC TGTGTAATCT	12780
CAGCACGTCC ACCATAGTTA AGAGCAAAAT TAAGAATCAA TCCTGTGTTG TTCTTAGTCA	12840
ATTCCTCAGC CTTGGTTAAA GCTTCAAAGG TTTGCTTAGG CAGGCGGTCT GTCTCCCCAA	12900
TCATTTGAAT CTTAACATTA TTCGCATGTA GTTCCGGGAC ATAATTATCA TAAAACTCTA	12960
CTGGCAAGTT CATGATAAAC TTGACTTCCT GATCTGGACG GGTCCAGTTT TCCGTAGAAA	13020
AAGCATAGAC CGTAATAACC TTGACGCCCA GTTTGTTGGC TGCCTTGGTC ACGGTTTGCA	13080
ATGCTTCCAT GCCCGCCTTA TGTCCAAAAA CTCGCGGTTG CATACGTTTT TTAGCCCAAC	13140
GGCCATTGCC ATCCATGATG ATGCCGATAT GAGCAGGAAC CTGTGTCGGA ACCTCTACTT	13200
CCACAGCCTT ATCTTTCTTA AAAAAATCAA ACATGATCTT ATTCCTATTC AAAAAATCTAT	13260
CGTTTCATTA TACCATATTT CCCCATTTTC TTCTATCACT AAGCTATTTA TTCTCAGGCA	13320
CCAAGCCCAT TTTTCAAAAA AATAAGCCGC CTGATTGGGC GACTTTATTT TTATAGGGAG	13380
ATTATTATGA AAAAGTTTAA GGAGTTTAAAG TTAAGGTCTT CTAACTTAT GAACCTTAGTG	13440
TACACTCCCT AGCTTAAAGT TTCCTTAAAGT ATTTTTAAAA ATCAAATTTT TCCATTTCCT	13500

616

CTGCCAATT	TTCTTGGATA	AACGTGTTTG	ATAGAGTTCC	ATTCGGTCTT	CATTTTCTAA	13560
GAAATGAGGA	GTTGGACGAA	CTTGAAAATT	CAAAATATCC	TCCAAACCAT	AAGGTACATA	13620
GAGTTCAAAA	TCTAATTCTT	CATTCAAGCG	CAGTCCAAC	GCCGTACACC	GTCTTGGATA	13680
CTTACTCATA	GCATCACGAG	AACTGGTATA	GGAAGCAGTG	TGAGGACTGT	GCTGATGCAT	13740
ATAGACCTGA	TTTTTCAATT	CCCACTGGTA	CTGAGGAAAA	TCCTCTCTCA	GCTTTTTCTC	13800
CAGTAATAAG	GTTTCCTCAT	AAGAAAAATC	TGGATCAAAG	AAAATCACAT	CTATATCTGT	13860
TTCATGATCA	AAAGGGGATT	TGTCTGACAA	AAGATTCCAG	ATGAAATTTC	TGACAGAACC	13920
TGCTGCCAAC	CACGAGTCTT	TCAAACCAAG	GTCTCGGATG	ATCGTCAGAA	TGGCCATCAT	13980
ATCTGGACTT	TCTCTAAAAG	CCTCTAAGAT	TTCTTGCTTA	TTTTTCACTG	TATTCATAAC	14040
CTAAGTGCTC	ATATGCCTTA	GCAGTCGCCA	CCCGTCCAGA	CCGTGTCCGC	ATGATAAAAC	14100
CTTTTTGAAT	CAAGTAAGGC	TCATACATGT	CTTCAACTGT	CTCACGCTCT	TCGGCGATAT	14160
TCACAGAAAG	AGTTCCTAGA	CCAACAGGTC	CTCCACTGTA	CATCTCAATC	ATGGTGCGAA	14220
GGATTTTTTG	ATCCACATAG	TCCAAACCTT	CATGGTCAAC	ATCCAGCATA	GTCAAAGCCT	14280
TATCGGTAAT	AACATCATCG	ATAACCCCAT	TCCCCATTAT	CTGGGCAAAA	TCGCGCACGC	14340
GCTTGAGGAG	ACGATTGGCA	ATACGAGGGG	TTCCACGACT	ACGTAGGGCC	AACTCAGATG	14400
CTGCCTCATG	GGTGATTTC	ATCTCAAAAA	TATCTGCCGT	CCGCTCGACA	ATTTCTGTCA	14460
AGTCAGCATG	AGCATAATAC	TCCATATGAC	CTGTAATCCC	AAAACGTGCC	CGTAGTGGAT	14520
TTGAGAGCAT	ACCAGCCCGA	GTCGTGCGAC	CAATCAAGGT	AAAAGGAGGC	AACTCCAAAT	14580
GAACACTGCG	ACTGCCTTCA	CCAGCCCCAA	TCATAATATC	GATGTAGAAG	TCCTCCATGG	14640
CACTATAAAG	CACCTCTTCC	ACTGACATGG	GTAAGCGATG	AATCTCGTCA	ATAAAGAGGA	14700
CATCTCCAGG	CTCTAAATCA	TTCAAAATCG	CTACCAAATC	ACCCGCTTTT	TCGATAACAG	14760
GACCAGACGT	TTGCTTGAGA	TTGACTCCCA	GTTCAATTGGC	AATGACAAAA	GCCATGGTTG	14820
TTTTCCCAAG	CCCTGGAGGG	CCAAATAAGA	GCACATGATC	CAGCGCTTCA	TCCCGCATTT	14880
TAGCGGCTTC	GATAAAGATC	TGAAGTTGAT	CCTTAACCTT	ATCCTGACCA	ATATATTAC	14940
GTAAATACTG	AGGACGGAGC	GTGCGTTCTA	CTAACTCCTC	ATCACCCATC	ATCTCATTAT	15000
CTAAAAATTCT	ACTCATGGCT	CTATTATATC	AAAAAAAACA	AGCCACAAAC	AAAAAAGCCA	15060
CCTGATTGGG	TGACTCCTAA	GTTTAGCACT	TATGTGGTAT	AATATTATAC	GGCACTTCTA	15120
CACCGCTTAC	GAAAGGAGGT	GAGATAGCCC	ATGATGGAAT	TAGTACTCAA	AACTATTATC	15180
GGACCAATTG	TGGTCGGTGT	CGTTCTTCGT	ATAGTCGATA	AATGGCTAAA	CAAGGACAAA	15240
TAGTGTCAAA	AAAGACCTCA	AGCTTATTTG	GTCGTGAGCT	TGGGGTCTTT	TCTAGCCTAT	15300

617

GATATAGAAC TAGTACTCAA TTCCTTTTTA TTATCCCATTA GTTCACGAAT TTTGTCAAAA	15360
CTTTACATTT TCTTCAACCG CTGTACGACA AGACGGTTAA GATTAAAGAGA ACGTTAGGGA	15420
TTCTATCAAT TTCATAGAAA TTTTGATTTC GTAAACGAAG AGACAATCTT ACATGTCACT	15480
TCTCATTTAA TACGCCACTA CTAGACAAGC AAAATCATT A TTACAGTAGT TCCAGTCCTT	15540
CAATTAACAG TCACTTACAA TCAAATTGAG TTTGAACTAG CTGAAGCGAC CACAGACCTA	15600
TTTCTTAGTC ATATTGCTA AAAAAATCCC CGCCAAAATC TCAAAAAGTC CCCGCCAATT	15660
CCCCGACCAA AATCCGAAAA ATACCGAAAA ATATCGAAAA ATTATTTTTA GAATAGTCCC	15720
AAAAATCCTG AAATAGAGCT AAAAACTCC ACCTGATTTC GTGGAGTTAA GGGAGATTAT	15780
TATGAAAAAG AAAAGTTAG GATTTTATTA AATAAAGTTA GGAGGTCTTT ATTTAATAAC	15840
TACATGATAC AAGACGAAAC TTAAACTAG CTTAACTTTT CTAAAAATTT ACTATTTTGC	15900
AAAAAATTC TATCACCAGC ACCTCACCAA TCGAGTAGGG GATAATCTCT AGCCCTCTC	15960
ACACCACCGT ACGTGCCGTT TGGCATACGG CGGTTCAACT AACTTTTAAC GCATGTCGTT	16020
CAAGGTAATA ATCCAAACAC GAAACCAGTC CACGTTTTTC CAGGACTGGT TTTGATATAG	16080
CACGTTTAAG TACCGACTTC TGAGCTACTA ATTGATAATG GTCGCCCCAG CCAGATACCT	16140
TATCTGCTAT CCATTTAGGA ACTCCTAACT TAAGCAATCC CCATAATCGT CTCGATTTCT	16200
TCTTCCATTG CTTCCAGATA ATCACTCGTA GGCGAGTACG CAAGCGCTCA TCTATGCTGG	16260
CGACTATACT TTTGATATTT CCCAATGAGC AATAGTTTAT CCATCCTCGA ATAGACAAAT	16320
TCAGTTGCTC AATACGTCTT GTTAGGTCTA TACTCCATTT CCTCTGTGTT AGTTTCTTCA	16380
ATTTAACTT AAATCTCCGA ACACTATCTT GATGTGGACG GCTTTTCCAA CCATCTGATA	16440
ATTTCCAGAA CCCAAACCT AGATATTTCA ACTCTCTTGG TCATGTTTAC TTTCAAACCT	16500
AGCCGTTTCT CAATAACGA CTGACTGAAT ACATC	16535

## (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8136 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

CCAGAGCGTT GCGTCCGAAA GTCTATCCAG ACACGGCTCT TAAAAACAA AAGGAGAAAT	60
GATGCATACT TATTTGCAAA AGAAAATTGA AAATATCAAA ACAACCCTAG GTGAAATGTC	120

618

AGGTGGTTAC CGTCGTATGG TTGCGGCTAT GGCTGATTTA GGATTTTCAG GAACTATGAA	180
GGCTATCTGG GATGACCTCT TTGCCCATCG TAGTTTGGCC CAGTGGATTT ATTTGCTGGT	240
TTTAGGAAGT TTTCCTCTCT GGCTGGAGTT GGTTTACGAA CATCGTATTG TTGACTGGAT	300
TGGGATGATT TGTAGCTTGA CAGGGATTAT CTGTGTAATC TTTGTATCGG AAGGTCGAGC	360
AAGTAATTAT CTTTTTGGCT TGATTAACTC TGTTATTTAC CTTATTTTGG CCCTACAGAA	420
AGGCTTTTAT GGTGAGGTGC TGACGACACT TTACTTCACA GTCATGCAGC CAATTGGACT	480
TCTAGTTTGG ATTTATCAGG CACAGTTTAA GAAGGAAAAG CAGGAGTTTG TCGCGCGTAA	540
ACTGGACGGC AAGGGCTGGA CAAAGTATCT TTCCATTAGT GTGCTTTGGT GGTGGCCTT	600
TGGCTTCATT TATCAGTCTA TTGGTGCCAA TCGTCCCTAT CGTGATTCAA TCACAGATGC	660
AACCAATGGG GTAGGGCAAA TCCTCATGAC AGCTGTTTAC CGTGAACAGT GGATATTCTG	720
GGCGGCTACC AATGTCTTTT CAATCTATCT CTGGTGGGGA GAAAGCCTGC AAATTCAAGG	780
GAAATATCTA ATTTATCTCA TTAACAGTCT AGTTGGTTGG TATCAATGGA GCAAGGCAGC	840
TAAGCAGAAT ACTGATTTAC TTAAGTAGGA AAAGATGTTT GAAAGTGCTG TTTTGAGATT	900
TCGATTAAAA CAGATATAGT TGATAATCAA GGATTTATAG TATGAAAAAG AGGATCGGCG	960
GGTCCTCTTT TGTTGTTGAA AAGATAAAAA ACTCAGTAAC CTAGAAATAA GACAACTGAA	1020
GCTTTACTCT ATATTCAATT TTTAGGAATG AGAAGGTCTA GATAAAATTG GACAACTTCC	1080
TGGTCTGTGA AATCTTGACC TTTTTTGAGC CACCAGGTCA ATGTCTCGAT AAAGTTGGAC	1140
ATGACCAAGT GTTGGAGGTA AGAAGTAGGC AGATTAGGGT GGGCTTCTTT TAAATTATCA	1200
GCTAGCACGG AATAGACATG GTGTTCTAGC TCTTTATGGA GTTGACGGAG GAAGTAGTCA	1260
TTTTTGGAAG ATAGCAGACT GGTGATATGG TCTTGGTTTT TATGAAAATG GAGAAAGAGG	1320
TGGGCGAGGT AGTCCTCGGT TGAAATGGCT TGCTCTCTTT CAAAAAGATG ATGGAAGAGG	1380
TAGCGGCAGA GCTGGTCCAG AAGAAGCTCC TTAAGTCTCAT AGTGACAGTA AAAGGTGGAT	1440
CGTCCACAT CTGCGAGATC AATGATATCC TGAACAGTAG TGGCCTCGTA GCCCTTAGCA	1500
TTCAAAAGTT GTATAAAAGC TTGATAGATG GCTTTTTTGG TTTTGCTGAT ACGGCGGTCA	1560
ATGTTAGTCA TATGGACACT TAAGGCAAAT TGTTCAGAAC TGAATAAAGC TGACGTTTTG	1620
CTTCTATCCT TTCTTTGAGT TTTAGTGGAT AATGATAATG AACAAGGTGT TCATAAATCT	1680
ATTATAACAA AGGAATGAGA AATATGAAGG CAAAATATGC TGTTTGGGTG GCTTTTTTCT	1740
TAAATTTGAC TTATGCCATT GTTGAGTTTA TTGCAGGTGG AGTATTTGGT TCTAGCGCTG	1800
TTCTTGCTGA CTCTGTGCAT GACTTGGGAG ATGCGATTGC AATTGGAATA TCAGCTTTTC	1860
TAGAAACAAT CTCCAATCGT GAAGAAGACA ATCAGTACAC CTTGGGCTAT AAGCGGTTTA	1920



619

GCCTGCTAGG AGCCTTGGTA ACAGCTGTGA TTCTCGTAAC GGCCTCTGTT CTAGTCATTT	1980
TGGAAAAATGT CACGAAGATT TTGCATCCGC AACCAGTCAA TGATGAGGGG ATTCTCTGGT	2040
TAGGAATTAT TCGGATTACT ATCAATCTGT TAGCGACTCT GGTGGTTGCT AAGGGAAAGA	2100
CAAAGAATGA GTCTATTCTG AGTCTGCATT TTCTGGAAGA TACGCTAGGG TGGGTAGCTG	2160
TTATCCTGAT GCGGATTGTT CTTTCGATTTA CGGACTGGTA TATCCTAGAT CCTCTTTTGT	2220
CCCTTGTCAT TTCTTTCTTT ATTCTTTCAA AAGCCCTTCC ACGTTTGTGG TCTACACTCA	2280
AGATTTTCTT GGATGCTGTG CCAGAAGGTC TTGATATCAA GCAAGTAAAG AGTGGCCTGG	2340
AGCGATTGGA CAATGTGGCC AGCCTTAATC AGCTTAATCT CTGGACTATG GATGCTTTGG	2400
AAAAAATGC CATTGTCCAT GTTTGTCTAA AAGAAATGGA ACATATGGAA ACTTGTAAG	2460
AGTCTATTCTG AATTTTCTTA AAAGATTGTG GTTTTCAAAA TATTACCATT GAAATTGATG	2520
CTGACCTAGA AACTCACCAA ACCCATAAGC GAAAGGTGTG TGACTTGGA CGGAGTTATG	2580
AGCATCAACA TTAGAAAAA GTGAAAAATA CTTGGGTACT ATCTTATTTG GAATAGAGTA	2640
ATTTCTTTAT TATTTAAATA TTTCAAAAAT TGGTAAGAGA AGAGCATTGT ATAACTCCA	2700
GATATATGAT TGTTAATGAT AAAAAATTTT CGATTAGATA CAAAATGCTT GACTTGGAGT	2760
CAACTCAAAG TTATATAATA AGATAAGTGA GTTAGAATAG CGTGAATTCA GTGAATGAAA	2820
TGAGAGGAGG TTAGCGTGTG AATATTAAAT CTGCCAGTGA TTTGTTGGGA ATTTCAGCGG	2880
ATACGATTCTG GTATTATGAA CGGGTTGGTC TTGTGCCACC GATTACTCGT ACTGCTACTG	2940
GGATTCTGTA TTTTCAAGAT CAGGATATCG AAGCGCTGGA ATTTATTAAG TGTTTTCTGT	3000
CGGCGGGTGT CTCTGTAGAT AGTTTAGTTG ACTATATGTC GCTCTACCAA AAGGGAGATG	3060
AAACGAGAGA GGAGAGGCTT GGTATTTTAC AAGAGGAAAA GCAAAAATTA GAGGAGCGCT	3120
TGTCTCAGCT ACAGACAGCT TTAAATCGTT TAAATCTCAA AATTAACTT TATAAGGAAG	3180
GAAAATTTTA AATGAAATCA GCAGTATATA CAAAGGCAGG TCAGGTTGGA CTTGCTAGCA	3240
TTGAACGTCC GCAAATAATA GAAGCGGATG ATGTGATTAT TCGTGTGCTT CGTGCGTGCG	3300
TTTGTGGTTC AGATTTATGG AGGTACCGTA ATCCAGAAAC GAAAGCTGGA CACAAAAATA	3360
GTGGACACGA AGCGATTGGG ATTGTTGAAG AAGCTGGGGA AGCCATTACG ACGGTGAAAG	3420
CAGGTGATTT TGTGATTGTC CCTTTTACAC ATGGATGTGG TGAGTGTGAT GCCTGTCTTG	3480
CTGGATTTGA CGGTCTTTC GACAATCATA TTGGCAATAA TTTGGGGGGT GATTTTCAGG	3540
CAGAAATATAT TCGCTTCCAC TATGCAAACT GGGCGCTGGT TAAAATCCCT CGTCAACCTT	3600
CTGACTATAC AGAAGGGATG CTCAAGTCCC TTTTGACTCT TGCAGATGTC ATGCCGACAG	3660

620

GCTATCATGC	GGCGCGTGT	GCAAATGTTC	AAAAAGGGGA	CAAGGTTGTT	GTTATCGGTG	3720
ATGGGGCTGT	TGGTCAATGT	GCTGTCATCG	CGGCTAAGAT	GCGTGGAGCA	TCACAAATTA	3780
TCCTTATGAG	CCGTCATGAA	GACCGTCAAA	AGATGGCTAT	GGAGTCAGGT	GCGACAGcTG	3840
TTGTTGCAGA	ACGTGGTCAA	GAAGGAATTA	CCAAGGTGCG	TGAAATCCTC	GGTGGAGGAG	3900
CAGATGCAGC	ACTTGAATGT	GTTGGTACGG	AGGCTGCTAT	AGAACAGGCG	CTAGGTGTTC	3960
TTCATAATGG	AGGGCGTATG	GGCTTTGTAG	GAGTCCCACA	CTATAATAAT	CGTGCTCTTG	4020
GTTTCGACATT	TATGCAAAAT	ATCTCTGTAG	CAGGTGGGGC	AGCTTCTGCT	ACAACATACC	4080
ATAAGCAATT	TTTACTAAAA	GCCGTCCCTG	ATGGTGATAT	CAATCCAGGT	CGCGTCTTTA	4140
CTTCAAGTTA	TAAACTGGAA	GATATCGACC	AAGCCTATAA	AGATATGGAT	GAACGTAAGA	4200
CAATTAAGTC	TATGATTGTA	ATCGAATAAA	AAACGAATAG	GAGTTTTAGA	ACTCTATTTCG	4260
TTTTTTATGT	TATCCTATTC	TTGATTTAGG	GTACTTTCTC	TTAATGTCAG	TCTGGTTCCC	4320
AGCATGGTCA	GGCTAGGGAT	TTTCCGACCG	TGGAGGACTT	CCTTGTTAAG	AATATCCATA	4380
CCTGCTCGGC	CCATTTCTTC	AGTATAAACT	GTAATACTAG	AGAGGGGAGG	ATAGACCTGT	4440
TTGGTCAGAC	TAGTGTCTGT	AAAGGAAATG	AGGCTGACGC	GATCTGGCAG	GCTGATTCCA	4500
GCTTCTTGGA	GGGCACGGAG	GGCACCGATA	GCTAAACTAT	CGCTGGCTGC	GAAAAATGCT	4560
GGCGGAAGTT	GGTCTCCCAA	GCTCTGAATG	GCCTCCTTCA	TTAAGTCATA	GCCAGACTGG	4620
GCAGTAAATC	TTCTTTGAAA	GACCAGTTCA	TCATGATAGA	TTCCCCTCGC	TTGACTATAG	4680
TTTTTGAAGT	TTTCTAGACG	CTTGCTCTGA	ATGATTICTT	CTTGGTCTGT	TGTTTCTTCA	4740
AGGCCTGTTA	GAATCCCGAT	ACGGTCCATT	CCTTGACTGA	GGAAATAATC	GACAACTGT	4800
TTCATAGCAG	TGTAAAAATC	CGTGATAATA	CAGGTATGTC	CCAGGGAAAG	TGTATCGCTG	4860
TCTAGAAATA	CAAGAGGCTT	TTGGTATTCT	TCAAAGGCAG	AAATCTGAGC	TCGACTAAAC	4920
TTTCCGATGC	AGAGAATCCC	AATCACTTCC	TCGCTTAGGG	TAAAAGGGTG	GTCATTAAAA	4980
TAGCGCAAGA	TATCATAGTC	CAACTCTTGG	GCTCTTTTTT	CTATTCTTAG	GCGAATCTGG	5040
TAGTAGTAGA	GGTCGTCCAG	CTCCCCTTGT	TCGCTGACCC	ATTGGATAAT	GGCAATCTTT	5100
TGCTTGGGTT	TGTGGGACTC	GCCTGTCTTG	AGGTGCTTGG	TGTAGCCCAG	CTCTTCAGCA	5160
ACGGTTAAAA	TACGGTGTCT	GGTTTCTTCT	GTAACAGATA	GGCTCTGGTC	GCGGTTGAGG	5220
ACGCGGGATA	CGGTCGCGAT	AGAGACAGAG	GCTAGCTGTG	CAATGTCTTT	TAAGGTAGCC	5280
ATAAATCCTC	CTTGATTAGG	TTAGTATATC	ATGTTTTTCT	TCTTTTTACT	GATATTTTAC	5340
TAAAAATTTA	GTAAAAAGGA	TTGACCTTGG	AAAATTCCTT	GGATATAATA	GAAAGAAAAAC	5400
GATTACACGT	TAAGATGGCT	TAACGGACAG	TCAAAGGAGA	ATTCATATGG	CACAACATCT	5460

621

TACTACTGAA GCCCTTCGCA AAGACTTTCT TGCTGTTTTT GGTCAAGAAG CAGATCAAAC	5520
CTTCTTTTCA CCAGGCCGCA TTAATTTGAT TGGTGAACAC ACAGACTACA ACGGTGGGCA	5580
CGTTTTTCCT GCTGCTATTT CCTTGGGAAC TTACGGTGCA GCTCGTAAGC GTGACCACCA	5640
AGTCTTGCCT TTCTACTCAG CTAACCTTGA GGACAAGGGC ATTATCGAAG TGCCTCTCGC	5700
TGACCTCAAG TTTGAAAAAG AGCACAACTG GACCAATTAT CCAAAAGGTG TCCTTCATTT	5760
CTTGCAAGAA GCTGGGCACG TGATTGACAA AGGTTTTGAT TTTTATGTTT ATGGAAATAT	5820
TCCAAATGGT GCTGGCTTGT CTCTTCTGCG ATCCTTGGA CTCTTGACAG GAGTCGTGGC	5880
TGAGCATCTC TTTGATTTAA AATTAGAGCG TCTCGATTTG GTTAAAATCG GCAAACAAAC	5940
AGAAAACAAC TTTATCGGAG TAAACTCTGG CATTATGGAC CAGTTTGCTA TTGGTATGGG	6000
GGCAGACCAA CGTGCTATTT ACCTAGATAC TAATACTTTA GAATACGACT TGGTGCCACT	6060
TGATTTGAAG GACAATGTCG TTGTTATCAT GAACACCAAC AAACGCCGTG AATTGGCGGA	6120
CTCTAAATAC AATGAACGTC GTGCTGAGTG TGAAAAAGCA GTGGAAGAAT TGCAAGTTTC	6180
CTTGATATTT CAGACTCTGG GTGAATTGGA CGAGTGGGCC GTTGACCAAT ATAGCTATCT	6240
GATTAAAGAT GAAAATCGTT TGAAACGTGC TCGCCATGCT GTGCTTGAAA ACCAACGTAC	6300
CCTCAAAGCT CAAGTAGCAC TCCAAGCAGG AGATTTGGA ACATTGAGC GCTTGATGAA	6360
TGCGTCACAC GTTCTCTGG AGCATGATTA TGAAGTAACT GGTGGAAT TGGATACCCT	6420
TGTTACACAC GCTTGGGCAC AAGAAGGAGT TCTCGGTGCT CGTATGACAG GGGCTGGTTT	6480
TGGTGGCTGT GGCATTGCCT TGGTTCAAAA AGATACTGTT GAGGCCTTTA AGGAAGCTGT	6540
AGGCAAACAC TACGAGGAAG TAGTTGGATA CGCTCCAAGC TTCTATATCG CTGAAGTTGC	6600
AGGTGGCACT CGCGTCCTTG ACTAGTCAAA AGGAGGCTCT ATAGTGACCT TAGTAAATAA	6660
ATTTGTAACA CATGTCATTT CTGAAAGCTC ATTTGAGGAA ATGGATCGAA TCTATCTGAC	6720
CAATCGTGTT TTGGCACGAG TGGGAGAAGG TGTTTTGGAA GTTGAGACCA ATCTGGATAA	6780
ATTGATTGAC CTCAAGGACC AGCTGGTTGA AGAAGCCGTT CGATTAGAGA CGATTGAGGA	6840
TAGTCAGACT GCGCGTGAAA TCCTTCGTGC TGAAGTATG GATTTGGTGA CTCCTTGTC	6900
AAGTCAGGTC AATCGTGATT TTTGGGCAAC CTACGCCAC TCTCCAGAAC AAGCGATAGA	6960
GGATTTTTAC CAACTCAGTC AGAAAAATGA CTACATCAA CTCAAGGCCA TTGCTAGAAA	7020
TATCGCTTAT CGTGTTCCAT CTGACTACGG AGAAGTTGAA ATTACCATCA ATCTCTCTAA	7080
GCCTGAAAAA GATCCCAAAG AGATTGTGGC AGCCAAGTTG GTGCAAGCTA GTAATTATCC	7140
TCAGTGTCAG CTTTGTCTAG AGAATGAGGG CTACCATGGT CGAGTTAACC ACCCAGCTCG	7200

622

TAGCAATCAC CGTATTATCC GTTTTGAAAT GGTGGGTCAG GAATGGGGTT TCCAGTATTC	7260
GCCCTATGCT TACTTTAATG AGCATTGTAT CTTTTTAGAT GGCCAGCATC GTCCCATGGC	7320
CATTAGTCGT CAGAGTTTTG AACGTCTGTT GGCTATCGTA GACCAGTTTC CAGGATATTT	7380
TGCTGGATCT AATGCCGACC TGCCGATTGT GGGGGGCTCT ATTCTAACTC ATGATCATTA	7440
TCAGGGAGGC CGTCACGTAT TTCCTATGGA ATTGGCTCCC TTGCAAAAGG CTTCCGATT	7500
TGCTGGTTTT GAGCAGGTCA AGGCTGGAAT TGTCAAGTGG CCCATGTCTG TCCTACGTTT	7560
GACTTCGGAT TCCAAAGAGG ATTTGATCAA TTTGGCTGAT AAGATTTTGC AGGAATGGCG	7620
CCAGTATTC AATCCTGCAG TGCAGATTTT GGCAGAGACA GACAGGACAC CGCATCACAC	7680
TATCACACCC ATTGCCCGCA AACGCGATGG ACAGTTTGAG TTGGACTTGG TCTTGCGAGA	7740
CAATCAGACT TCAGCAGAGT ATCCTGATGG TATCTATCAT CCCCACAAGG ATGTCCAACA	7800
TATCAAGAAG GAAAATATCG GCTTGATTGA GGTCTATGGC TTGGCAATCT TGCCACCACG	7860
TCTGAAAGAA GAAGTGGAGC AAGTCGCTAG CTATCTTGTG GGAGAAGCTG TTACAGTTGC	7920
CGATTATCAT CAGGAGTGGG CAGACCAACT CAAATCCCAA CATCCAGACT AACGGATAAA	7980
GAAAAAGCCC TTGCAATCGT CAAGGACTCT GTGGGTGCTA TCTTTGCGCG TGTACTTGAG	8040
GATGCAGGAG TCTACAAGCA GACAGAACAA GGCAGACAG CCTTTATGCG CTTTGTGGAA	8100
CAGGTCGGAA TTTTACTAGA CTAGGAGCTT TCTCGG	8136

(2) INFORMATION FOR SEQ ID NO: 76:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10011 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

CCCATAGTGA AGAGTGGCCA TAAGAAGGTC TTCTAGGCTT AATTTAGGTT TTCGTCCACC	60
TTTTGCGTGT TTAAGTTGAT AAGCTGTTTT TAACACAGCT GAACATCTCT TCAAAAGTCG	120
TGCGCTGAAC ACCAACAAGA CATTAAATC GTGTATCAGT TAGTTGTTTA CTTGCTTCAT	180
CATTATAGA ACTACTATAC CATGTTTTGT TTCGCAGGAA GTCTAATATT GTCAAATACT	240
GGAACGCTCA TTGCTGGGAT ACGGAATAAG ATTGGCCCAG CTTTCGATAAC TGGGATACCT	300
GGTTCAAAAC CAAGGTCTGT TGCAGCGATT GGTGTAAAGA TATCGTAACC TTTCATAAGG	360
TCTTCGTTTA CATCTTTCAC CATAACTGCA TCACAGTGAA CATCGTAACC ACGGTTTGAA	420
AGTTCTTCTT CTAGAGCACT TTTAATTTGG TGAATTGAGT TAACACCTGC ACCGCAGGCA	480

623

GCAAGAATTT TAATCATTTG GATTTCTCC GATTTTATTT TTTAATAGAC AAGATTAAGC	540
GGTTGCTTCA GCAATGTAAG CATAAAGGGC TTCTGGTTCA GAAATTTTTC ATAGGTCTTC	600
AAGATGACCA TTCTCTGTGA AGAAGTCCAT TAAC TGAGCA AGAATGTTTCG TTTGACTTGA	660
ACTTGAATTA TTGATGATAA AGAAGAGCAA GGATACTTCT ACTTCCTTAC CTGGCGCAAT	720
CATATTATGG AAAGTCACCG GTTCTCTAA TCGAACAACC ACCACTTTCT CAGCTAGATT	780
ATGAACAATA TCTGTGTGAG GAATCATTAC ATTTGCAAGT CCTTTCCTAG AAATTCCATA	840
TATAAACCAAG TTGGAAATGA CTTTTCACGC GTGATCAAGG CTTACAGATA AGTTGGAGTG	900
ACAATTTCTC GTTCTTCCAA CAAGCTTGCT ACCTGATCAA AAAGTTATTC TTGATTATCC	960
GCTTCTAAGC AAAACACAAG GTTTTTGTCA AAGAAATAAT CTAATACCAT AAGGTTTTTC	1020
CTTCTTTCCA TTAACCTTAT GCTATAAGTA TAACACTATA TGAAATCGTT GTTAATTACT	1080
TTCTATTCTT TTTTGTCTCT TTTTATATAT TTTGTTTTG TTTATAGTTT GTTATATAAA	1140
AATAACACA CAAACAAATA CTCCAAGCAT TTTCTGTTC TAATACTCAA TGAAATCAA	1200
AGAGCAAAC AGGAAGCTAG CCGCAGTTGT TCAAAACACA GTTTTGAGGT TGTAGATGAA	1260
ACTGACGAAG TCACTCAAAA CATGGTTTTG AGGTTGTAGA TGAAACTGAC GAAGCAACAG	1320
CCATACATAC GGTAAGGCGA CGCTGACGTG GTTTGAAGAG ATTTTCGAAG AGTATAAAAA	1380
CTAAAAAGC AGACCATCTA AGCCTGCTTT ACTATTGATT CTTATATAAA TTTCTGTGA	1440
ACAAGGAAAG GCATTTCTGA TAACTTATTC TTCATCCATA CTCAAGACGC TGAGGAAGGC	1500
TTCTTGCGGA ACTTCAACTG ATCCGATGGA TTTCATGCGT TTCTTACCAG CTTTTGTTT	1560
TTCAAGGAGT TTACGCTTAC GAGAAACGTC ACCACCATAA CATTTAGCAA GTACGTTCTT	1620
ACGAAGGGCC TTGATATCAG TACGAGCGAC AATCTTGTGT CCAATAGCCG CTTGGATTGG	1680
AACTTCAAAT TGTTGGCGAG GGATGATTTT CTTGAGTTTA TCAACGATGA GTTTCCCACC	1740
TTCTAGGCA AAGTCCTTGT GAACGATAAA GCTGAGGGCA TCCACCTTAT CTCCATTGAG	1800
AAGAATATCC ATTTTCACCA GCTTAGATGG GCGATATTCT GACAATTCGT AGTCAAAGCT	1860
TGCATAACCA CGTGTCGAAG ACTTAAGTTT ATCAAAGAAG TCAAAGACAA TTTCAGCAAG	1920
AGGAATTTGA TAGATAACAT TGACACGGTT ATCATCAATA TAGTCCATAG TCACAAAGTC	1980
CCCACGCTTA CGCTGAGCTA GCTCCATTAC TGCTCCGACG AACTCCTGTG GTACCATGAT	2040
TTGCGCCTTG ACATAAGGCT CTTCAATGGT CGCAATCTTA GTTGGGTCTG GAAACTCAGA	2100
TGGGTTAGAC ACATCCATAG ACTCACCGTC GGTCAAATTA ACTTTGTAAG TAACAGACGG	2160
AGCTGTCATG ATGAGGTCAA TATTGAACTC ACGCTCTAAA CGTTCCTGGA TAACATCCAT	2220

624

ATGGAGAAGT CCAAGAAATC CACAACGGAA ACCAAATCCA AGTGCCTGAG ATGTTTCTGG	2280
TTCAAACCTGA AGACTAGCAT CATTCAGTTG CAATTTTTC AAGGCTTCAC GCAGGTCATT	2340
GTACTTGTTT GATTCGATTG GGTAGAGACC CGCAAAGACC ATAGGATTCA TCTGCTTATA	2400
ACCATGTAAT GGTTCGCGG CAGGATTGGT TGCCAAGGTA ACGGTATCAC CCACACGAGT	2460
ATCCTGAACC GTCTTGATAG ACGCCGCAAT GTAACCAACA TCACCAGTCG CAAGGAAATC	2520
ACGACCAACC GCTTTTGGTG TAAAAATACC GACTTCGGCC ACATCAAAGG TCTTACTATT	2580
GCTCATGAGC TGAATCTTAT CACCAGGTTT GACCACTCCG TCCATGACAC GCACTTGGAG	2640
GATAACCCCA CGGTAAGCAT CGTAAACAGA GTCGAAAATC AAGGCCTTAA GTGGCGCCGT	2700
CACATCACCC GTTGGTGCTG GTACTTTTTC TACAATTTGC TCGAGGATTT CTTCAATCCC	2760
AATACCAGCC TTGGCAGAAG CCAAACTGC TTCACTGGCA TCCAAACCAA TCACATCTTC	2820
AATCTCTGTA CGCACGCGCT CCGGATCTGC AGCCGGCAGG TCAATTTTAT TAATGATAGG	2880
CATGATTTCC AAATCATTAT CCAAAGCCAG ATAAACGTTG GCAAGAGTTT GAGCCTCAAT	2940
TCCTTGAGCC GCATCGACCA CAAAATAGC ACCCTCACAG GCAGCTAGCG AACGTGAAAC	3000
TTCATAGGTA AAGTCAACGT GCCCTGGTGT GTCAATCAAG TGGAAAATAT AAGTTTCCCC	3060
ATCTTTTGCA GTGTAATTCA ACTCGATGGC ATTCAACTTA ATAGTAATTC CACGTTCCCG	3120
CTCTAGCTCC ATGCTATCCA AAAGCTGGGC CTGCATTTCA CGACTTGAAA CCGTCTCTGT	3180
TTTTTCCAAA ATGCGGTCTG CTAGAGTTGA TTTTCCGTGG TCAATATGGG CGATAATAGA	3240
GAAGTTACGG ATCTTCTCCT GTCGTTTTTT CAATCTTCT AAGTTCATGA TTTCTTCTCT	3300
TTCAGGGTAT CTATTTATTA TAAATTGTTT TTGATATTTT GACAAGACCA TACCTGCTA	3360
GGAGTACTAA TCTTCAGCGA CAAAGCCGTC ATTTTCGATA AAGTGGTGT CTGTCAATTCC	3420
TTGGTCTGTA AAGACAATCC CGTGAAGGAC ACCACCATAA ACAGCTCCTC CATCCATTCC	3480
AATCTTGCCA TCTTCTGTAG TCCAAAGCTC AGATGTACCG CGTTCTTGCT GTAACAAACC	3540
ATAGACCGGT GTATGACCGA AGACAATGGT TTTTCCAGTA TGATTTTCAG CTCGGTGGAA	3600
TGGTTTTCTA AGCCATACTT TTTTATAATC TGTGTTTTCA TGCCAGTCGT CCAAGGTCAA	3660
ATCAATACCT GCGTGAACAA AGATATACTT GTCTGTCTCT ACTACAAATG GCATTTGACG	3720
AATGAATTCG ACCAAGTCTG CCGCTTCAGC GgCAACCCGC TTGGCATCTT CTACTCCATC	3780
AACTGGTGCA TCCAAGGGAC GACCTAGGAT AGAGTTAATG GTTGATCTC CACCATTGCG	3840
ACTATAATGG TCATAACTTT CTTCTGGGTC ATCTAGCCAA GTCAAAAACA TATACTCGTG	3900
GTTCCGGAC AAACAGATAG CCCCTTGATT GTCCACCAAG TCCTTGACCA TTTCAAGAAC	3960
ACGGTGAATA TCCTCACCTC TGTCATCAA ATCACCTAGA AAGAGCAACT GGGGCTGACC	4020

625

ATCCCAGGTT TTGAGAAGGT CTTCCAGCAT CCCAGCTTTT CCGTGAACAT CTCCAATTAC	4080
ATAATAATCT GTCATCTTAT TTCTCCCTGT TTCTCAACAA TTCTCTTGCT TCGTCAGGG	4140
CTGCTTCTGT CACATCATCA CCTGCCAACA TCTTGGCAAC TTCCTCCACT CGCTCTTCGA	4200
CCGTCAAGAG ACGAACAGTC GAAACCGTTG AATGGTCATT ACTAATCTTC TCAATAAAGA	4260
ATTGATAATC TGCAATCGCA ATTACTTGTG GCAAAATGGGA GATAGCCAAA ACCTGACCAT	4320
GCTGACCAAT TTTATGAATT TTCTGAGCAA TAGCTTGAGC AACACGACCT GAAACTCCCG	4380
TATCCACCTC ATCAAAGACA ATGCTAGTCT TGCCTTCTTT ACGTGAAAAG GCAGACTTAA	4440
TGGCTAACAT GAGACGAGAT AATTCCTCTC CAGAAGCAAC CTTAACCAAG GGTTTAAAGT	4500
CTTCTCCAGG GTTGGTTGAA ATATAAACT CAACCATTTT ATTTCCCTCA CGACTGAATT	4560
TTCCCTTACT AAAACGAACC TGAACTGGG CTTTTTCCAT ATAAAGATCT TGCAGTTCTT	4620
GTTTAATCTC AGCTTCGAGT TGCTGAGCCA AATTATGACG AGCAGAAGCA AGTTGACCTG	4680
CCAAATGAC AAGATTGACT TCCAACTTCT TAAGCTCTGC TTCCATGTCC TCAGACGAAA	4740
GATTATTGCC TGTCAAGAGA TTGTATTCTT CCGTAATCTT GGCAAAATAA AGCAAAACAT	4800
CATCAACAGT CCCACCATAC TTACGAGTAA TAGTATGAAG GAGGTCCAAA CGATTCTCAA	4860
CCTGCATCAG GCGATTGCCA TCAAAATCAA GGTCTCAAT GATAGCTTCC AAACGTTTGC	4920
TAATGTCTTC TAAAACATAG TAGGTCTCAG ACAGATAGCT TGAAATTTCA CGGTATTGAG	4980
GATCATACTC TTCGACACTT TCCATGTCTT TCATAGCTGA ACGAACATTG GCCAGACTTG	5040
AAAAATCTTC ATTGTCCAAC ATACTGTAGG CATTGGTCAG TGTATCCGCA ATATTTTGT	5100
GGTTGAGGAG TTTATCTCCG TCTTGATTGA GAGCCAAGTC TTCTCCAGCC TGCAAGTTTG	5160
CTGCCTCAAT CTCTGCCATT TGAAATTCCA ACATTTGAT ACGTGCCTTG TGTTCCTGTT	5220
GGTTTTTCTT GACTTCCAGA ACCTGCTTGC GCATTTTCCG ATAGGCATCA AAACCTGTTT	5280
GATAGGTTTC TTTCAAGTCC CAAAAGCGG CATCACCAAA TTCATCCAAC ATCTGGATAT	5340
GCAGTTGGGG ACGCATTAAAC TCCTCATGGT CATGCTGACC ATGAATATCT ACAAGATGTT	5400
GCCCAATAGC TCGCAAAACA GACAGATTAA CCATCTGACC ATTTACACGG CTGATACTAC	5460
GACCATTTTG CAAGATTTCC CGACGGATGA TAATTTTCATC ACCTAATTCT AAACCTTGCT	5520
CATCAAAAAT TTCCTGTAAA AGACGACTAT TCTCAACTGA GAAAAGCCCC TCAATCTCTG	5580
CCTTTGGTGC ACCATGACGA ATAACATCTG TCGTCGCACG AGCTCCCAAC ATCATATTCA	5640
TGGCATCAAT GATAATCGAC TTCCCTGCAC CCGTTTCACC AGTCAGGACA GTCATCCCT	5700
TTTCAAAATT GAGGGAAATA GCCTCAATAA TGGCAAAGTT TTTTATCGAA ATTTCAAGTA	5760

626

ACATATAGAC CTACCAATT TTTACTTGTT CAAAGATTTC CTCTGCTAGA CTTCCTACTTC	5820
TGGCAATGAC TAAAATCGAG CTATCATCAG TCAAACAGCT AAAAATCTTG TCTGCAAAAG	5880
TCTCGATTAA CTGAGCTTTT ACAAAGCCG TATTTCTGG AATAACTTGG AGATTGATCA	5940
TCTTATCCAT CAATTCAGCC GATTCGATAT TGTCTTCAGC CAGTTGCAGA CTTTTACGA	6000
TTGATTTTGG CAATTCGTAG ACATAGGTGT TGCTCTCAA AGGAATTTG ACAATACCTA	6060
ACTCTTTGAT ATCTCGGGAT ACCGTCGCCT GAGTGGCAGT GATACCTGCT TCTTCAAAT	6120
GTTCTACAAT TTCTTCTGC GTGCCGATTT GATAATCTGT CACCAATCTT CTAATTTTTT	6180
CAAGTCTCTC TTTTATTC ATTTTAAAT TGACTATGCG CCCTCTCTAC TGCTTCTTTA	6240
ATCTCAGCAA GAATCTGATT GCTTGCTGAC TTTCTTTTT TCAAATACGC TAAAAATCA	6300
ATATTTCCAT GTCCACCTTG GATGGGAGAA AAGTCCAAGC CAAGGACTGA AAAACCTACC	6360
TCTACTGCCA TAGCTGTAC AGATTCAAGG ACATTCTGAT GAACCTTAGC ATCTCGAATA	6420
ATTCCATTTT TCCCAATCTG CTCACGTCCT GCCTCAAAT GAGGTTTGAC AAGTGCTACC	6480
ACCTGACCTT GATCAGCCAA GACACGGTGC AAGGCTGGCA AAATCAGACT AAGGGAAATG	6540
AAACTCACAT CAATACTGGC AAAGCTCGGC TCCTGCTCGA AATCAGTCTT TTCAGCATAG	6600
CGGAAATGA ACTGCTCCAT GCTGACAACT CGTGGGTCTT GCGTAATTT CCAAGCCAAC	6660
TGATTGGTAC CAACATCGAC TGCAAAGACC AACTTGGCAC TATTCTGTAG CATGACATCG	6720
GTAAACCTC CAGTAGAGGC CCCGATATCA ATCGTAGTCG CGCCATCCAC CGACAAATCA	6780
AAGACCTGCA AGGCCTTTTC CAGTTTCAA CCACCAGGC TGACATACTT GAGTTTCTCC	6840
CCCTTGAGTT TTAATTCGGT GTCATCTGGA ATTTCTCTC CTGGCTGTG AAACCGTTCT	6900
CCATTAAGGA CTGCTACGAC TAGGCCAGCC ATCACACCTC GCTTGGCCTG CTCTCTCGTT	6960
TCAAACAACC CCTGTTTATA AGCTAGTACA TCCACTCTT CCTTAGCCAT TGATTCTCAA	7020
ACTTTCTACT ACACCTACAA TCGATTCTGT TCAAAGGGA AGCTGCTGGG CAATTTCTTC	7080
TAATTTTCA TTAGCTTGAT CCAGGGTTG GTTACAAAAG GCAATGGACT CTCCAAGCC	7140
CAACAGGGCA GGATAGGTG ATTTTCTGC CTGCAGATCC TTTTGAGTG TCTTGCCGAT	7200
TTCTCAAAA CTAGCTGTCA CATCCAGTAC ATCATCTCTG ACTTGAAAAG CAAGTCCAAT	7260
CAATTCACCC ACAGTTTCA GCTTCACCTG CATTCAGGT GACAATTCAG CTATAATAGC	7320
TGCCGCTTG AAGGGATAGG CTAGTAACCT CCCAGTCTTA TTGGCATGAA TAGTCTGAAG	7380
TTCTTCCAAA GACAAGTGCT GGTGTTGCCC TCCCATATCC AAAACTTGCC CTGCTACCAT	7440
ACCCAGACTA CCTGAAGCAA GGGATAAGTT GGCAATCAAG TCCACCTTAA TCTGACTTGG	7500
CAAACTGCCC TGCGCAATCA AGGCATATGA GTCTAAGAAT AAGGCATCTC CAGCCAAAAT	7560



627

GGCCATAGCT TCACCGAATT TCTTGTGATT GGTAAACCGC CCTCTTCGAT AATCGTCATC	7620
ATCCATAGCA GGAAGGTCAT CGTGAATCAA GCTCCCTGTA TGAATCATCT CTAAGGCAGT	7680
AGCTACCTGC GCGTGAGCAG GTTTGATGGT AACCTGCAAG GCTTCCAGAA CTTCTAACAA	7740
GAGAAAAGGC CGAATACGCT TGCCACCAGC ATGAATAGAA TAGAGAACAG ACTCCCGTAA	7800
ACTAGAGGCA AACTGCTGGT CTCCATAAAA ATCTTCCAAA GCCGACTCGA CAAGAGCTAA	7860
TTTTTCTTGC TTTTTCATTC AAAATCACCT TCTGTTCCGT CTTCTTGAT GACCTTGACC	7920
AAGGTCTTTT CAGCCTTGTC CAGCGTAGCT TGGAGCTCTT TTGACAAGAC CATGCCCTTT	7980
TGAAAGGCAG TAATCGCATC TTCCAGAGCA ATTCACCAT TTTCCAACT TTGGACAATG	8040
GTTCAGTCT CTGCTAGATT TTCTCAAAT TTCTTTTGT TTGACATCTT TAACCTCTAA	8100
TTCTACTTGA CCATCTCGCA TCAAAGCGT TACTTGGTCT TTTTCTTCA AACTCTCAAC	8160
CGAATCTACA ACGGACTCTT CTTTTTTGAC AATAGCATAA CCACGCGCCA CGATTGGCT	8220
AGTATCCAAC ATGAGCAAAG CTTCCGAAAG TCGCTTGGCC TCAGCAACCT TGGCGTCATA	8280
AACTAACGCC ATTTGGCTAC CTAAGAGCTT GTCCAACGT CCTAAACGGT CTTGATAGCG	8340
TTGGATTTTG GTAACAGGTG ATAATTGTAC TAATTGATGA GTTCTTGCTT GAACTAATTG	8400
TTTGTATCA GAAATCCGAG TTCCGAACT TTGTTTCAA CGCAGTTGCA GTTGGTCCAA	8460
GCGTTGCAAA TAACCGTCAT ACAAGCGCTC AGGTTGTCTA AAGATAACAG ACTGACTGCA	8520
TTTTTTCAA GCCTCTTGT TCTTAGATAG AACATTTCCG ACTGCCGTTA CCATCCGTTT	8580
TTCTGATTTT TGCAAATGAG CTAATACATC CAACTTGGTC ACAGGTGTTG CCAGTTCAGC	8640
CGCCGCTGTT GCGTTCGAG CGCGTCGATC TGCCACAAAA TCTGCCAAGG TCACATCCGT	8700
CTCATGCCCC AACTAGAGA TAACTGGCAA ACGAGATTCA AAAATAGCTC GTACCACAAT	8760
TTCTTCGTTA AAGGCCCAGA GATCCTCAAT AGAACCACCT CCACGACCAA TAATGAGCAA	8820
ATCCAAATCG TCCCGTTGAT TAGCACGCGC AATATTTCTA GCATTTCTT CCGCAGCCCC	8880
TTACCTTGA ACCTTGGTCG GATAAAGAAG GATGTCAACA CCTGGGAATC GCCTGCTGAC	8940
GGTCGTGATA ATATCTCGAA TAACGGCTCC ACTACGGCTG GTTACTACAC CAATTCTCTT	9000
AGAAAATTGG GGCAGAGCTT GCTTGAAGCG TTCTTGAAAC AGGCCTTCTT CTGTCAATTT	9060
TTTCTTAAGT TGTTCAACT GAATCGCAAG CGCCCCAACC CCATCAGGCT CAGCTTTTTC	9120
AATGATGATG GAGTAGCTAC CACTTGGTTC ATAGACCTGT ACACGCCCAA TCACATTGAT	9180
CTTCATTCTT TCTTCCAGGT CAAACCCTAA TTTCTGATAA ATCCCAGACC AGATGGTCGC	9240
TTGAATAACT GCATGGTCAT CCTTTAGGGA GAAATATTGG TGAGTAGGTC GTTTACGAAA	9300

628

GTTGGAAACT TGACCAGTTA AATAGACCCG TTCCAAGTAT GGGTCTTTAT CGAATTTTCAT	9360
TTTCAGATAC TTGGTCAAAG TTGTTACCGA TAAATACTTT TCCATCTCCA CCTACTATTTC	9420
ATTTACTTGC TCTTTCATGG GTATTATTAT ACCAAAAATA TGCCTAAAAA TCTCCATTTA	9480
TGTACCATTA TGAGGGAAAA ATAGAAAAAG GAGGCAAGGC CTCCACATGT GATTATTTGC	9540
TGTTTCGAGC TTCTTCCAAA ATCTTTGCAA TCTTGGTCGT CAACAGGTCG ATAGCCACGG	9600
TATTGCTAAC CCCTTCAGGA ATGACGATAT CAGCATAACG CTTAGTTGAC TCGATAAACT	9660
GGTGGTACAT TGGTTTGACC ACACCTAAGT ACTGGTTAAT AACGCTATCA AGGCTACGGC	9720
CACGCTCCTC CATATCACGC TTGATACGAC GAATAATGCG CACATCGTCA TCCGTATCCA	9780
CAAAAATCTT GATATCCATC AAATCGCGCA GACGCTTGTC CTCCAAGACC AAAATACCCT	9840
CAACGATAAA GACATCTTGA GGTTCCTGAC GATAGGTCTT GCTACTCCGT GTATGCTCTG	9900
TATAGTCGTA GGTCGGGATG TCCACCGGAC GCCCTGCCAA CAATTCCTTA ATCTGCTCGA	9960
TCATCAAGTC TGTATCAAAG GCAAAAGGAT GGTTCATAGTT GGTTTTGACG G	10011

(2) INFORMATION FOR SEQ ID NO: 77:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5365 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

CGTGTGGTCT TAAAAATAGA AGACAAAGAA CAACTGTTG GAGGCTTTGT CCTTGCAGGC	60
TCAGCCCAAG AAAAAACCAA AACAGCTCAA GTTGTGGCTA CTGGACAAGG TGTTCGTACC	120
TTGAACGGTG ACTTGGTTGC TCCAAGTGTT AAACTGGAG ATCGTGTCTT AGTTGAAGCC	180
CACGCAGGTC TTGATGTCAA AGATGGCGAT GAAAAGTACA TCATCGTAGG CGAcTAACAT	240
TTTGGCAATC ATTGAGGAAT AGAAGGAGAA AGTAAGTATG TCAAAAGAAA TTAAATTTTC	300
ATCAGATGCC CGTTCAGCCA TGGTTCGTGG TGTCGATATC CTTGCAGACA CTGTTAAAGT	360
AACCTTGGGA CCAAAAGGTC GCAATGTCGT TCTTGAAAAG TCATTCGGTT CACCCTTGAT	420
TACCAATGAC GGTGTGACCA TTGCCAAAGA AATCGAATTG GAAGACCATT TTGAAAATAT	480
GGGTGCTAAG TTAGTATCAG AAGTAGCTTC TAAACCAAT GATATCGCAG GTGACGGAAC	540
TACGACTGCA ACAGTCTTGA CCCAAGCTAT CGTCCGTGAA GGAATCAAAA ACGTCACAGC	600
AGGTGCAAAT CCAATCGGTA TTCGTCGTGG GATTGAAACA GCAGTTGCCC CAGCAGTTGA	660
AGCTTTGAAA AACAAACGCCA TCCCTGTTGC CAATAAGAA GCTATCGCTC AAGTTGCAGC	720

629

CGTATCTTCT CGTTCTGAAA AAGTTGGTGA GTACATCTCT GAAGCAATGG AAAAAGTTGG	780
CAAAGACGGT GTCATCACCA TCGAAGAGTC ACGTGGTATG GAAACAGAGC TTGAAGTCGT	840
AGAAGGAATG CAGTTTGACC GTGGTTACCT TTCACAGTAC ATGGTGACAG ATAGCGAAAA	900
AATGGTGGCT GACCTTGAAA ATCCGTACAT TTTGATTACA GACAAGAAAA TTCCAATAT	960
CCAAGAAATC TTGCCACTTT TGGAAAGCAT TCTCCAAAGC AATCGTCCAC TCTTGATTAT	1020
<hr/>	
TGCGGATGAT GTGGATGGCG AGGCTCTTCC AACTCTTGTT TTGAACAAGA TCGTGGAAC	1080
CTTCAACGTA GTAGCAGTCA AGGCACCTGG TTTGGTGAC CGTCGCAAAG CCATGCTTGA	1140
AGATATCGCC ATCTTAACAG GCGGAACAGT TATCACAGAA GACCTTGGTC TTGAGTTGAA	1200
AGATGCGACA ATTGAAGCTC TTGGTCAAGC AGCGAGAGTG ACCGTGGACA AAGATAGCAC	1260
GGTTATTGTA GAAGGTGCAG GAAATCCTGA AGCGATTTCT CACCGTGTG CGGTTATCAA	1320
GTCTCAAATC GAAACTACAA CTTCTGAATT TGACCGTGAA AAATTGCAAG AACGCTTGGC	1380
CAAATTGTCA GGTGGTGTAG CGGTTATTAA GGTTCGAGCC GCAACTGAAA CTGAGTTGAA	1440
AGAAATGAAA CTCCGCATTG AAGATGCCCT CAACGCTACT CGTGCAGCTG TTGAAGAAGG	1500
TATTGTTGCA GGTGGTGGAA CAGCTCTTGC CAATGTGATT CCAGCTGTG CTACCTTGG	1560
ATTGACAGGA GATGAAGCAA CAGGACGTAA TATTGTTCTC CGTGCTTTGG AAGAACCCGT	1620
TCGTCAAATT GCTCACAATG CAGGATTGTA AGGATCTATC GTTATCGATC GTTTGAAAA	1680
TGCTGAGCTT GGTATAGGAT TTAACGCAGC AACTGGCGAG TGGGTTAACA TGATTGATCA	1740
AGGTATCATT GATCCAGTTA AAGTGAGTCG TTCAGCCCTA CAAAATGCAG CATCTGTAGC	1800
CAGCTTGATT TTGACAACAG AAGCAGTCGT AGCCAATAAA CCAGAACCAG TAGCCCCAGC	1860
TCCAGCAATG GATCCAAGCA TGATGGGCGG GATGATGTAA GCTTTCTATA GAAAACAACT	1920
TATAAAAAAC ACAAAAGGAG GGAATGACTA ACCCTTCTTT TTATAGGCTC TTTGTCAACT	1980
GTAGTGGGT GAAGTCAGCT AAGCTCGAGA AAGGACAAAT TTCGTCCTTT CTTTTTTGAT	2040
GTTCAAAGCG ATAAAAATCC GTTTTTTGAA GTTTTCAAAG TTTCGAAAAC CAAAGGCATT	2100
GCGCTTGATA AGTTTGATGA GATTATTGGT CGCTTCCGGT TTGGCGTTAG AATAGTGATG	2160
TTGAAGGGCG TTGATAATCT TTTCTTTATC TTGAGGAAG GTTTTAAAGA CAGTCTGAAA	2220
AATAGGATGA ACTTGCTTAA GATTGTCTC AATAAGTCCG AAAAAATTCT CCGGTTCTT	2280
ATTCTGAAAG TGAAACAGCA AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTA	2340
ATAGCTCAAA AGCTTGCTTA AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGACC	2400
ATAAAATCGC TTATCACTCA GTTTACGGCT ATCTGTGTG ATGAGCTTCC ACTAGCGCTT	2460

530

GATAGCCTTG	TATTCATGGG	ATTTTCGATC	CAATTGGTTC	ATAATTTGAA	CACGCACACG	2520
ACTCATAGCA	CGGCTAAGAT	GTTGTACAAT	GTGAAAGCGA	TCCAACACGA	TTTtagcatt	2580
CGGGAGTGAA	ACAGTCTGGG	AGACTGTTTC	AGCCTGAGCC	TAGAAATTTG	AAAGCGAAGC	2640
TGTTTAGCCA	AGTCATAGTA	AGGACTAAAC	ATATCCATCG	TAATGATTTT	CACTTGACAA	2700
CGAACGGCTC	TATCGTAGCG	AAGAAAGTGA	TTTCGGATGA	CAGCTTGTGT	TCTGCCTTCA	2760
AGAACAGTGA	TAATATTAAG	ATTATCAAAA	TCTTGCAGAA	TGAAACTCAT	CTTTCCCTTA	2820
GTGAAGGCAT	ACTCATCCCA	AGACATAATC	TTTGGGAAGCC	GAGAAAAATC	ATGCTCAAAG	2880
TGAAAGTCAT	TGAGCTTGCG	AATGACAGTT	GAAGTTGAAA	TGGCCAGCTG	ATGGGCAATA	2940
TCAGTCATAG	AAATTTTTTC	AATTAACTTT	TGAGCAATCT	TTTGGTTGAT	GATACGAGGG	3000
ATTTGGTGAT	TTTTCTTTAC	CAGGGGAGTC	TCAGCAACCA	TCATTTTTGA	ACAGTGATAG	3060
CACTTGAAAC	GACGCTTTCT	AAGGAGAATT	CTAGAAGGCA	TACCAGTCGT	TTCAAGATAA	3120
GGAATTTTAG	AAGGTTTTTG	AAAGTCATAT	TTCTTCAATT	GGTTTCCGCA	CTCAGGGCAA	3180
GATGGGGCGT	CGTAGTCCAG	TTTGGCGATG	ATTTCTTGT	GTGTATCCTT	ATTGATGATG	3240
TCTAAAACT	GGATATTAGG	GTCTTTAATA	TCGAGCAGTT	TTGTGATAAA	ATGTAATTGT	3300
TCCATATGAA	TCTTTCTAAT	GAGTTGTTTT	GTGCTTTTC	ATTATAGGTC	ATATGGGACT	3360
TTTTTTCTAC	AACAAAATAG	GCTCCATAAT	ATCTATAAGG	GATTTACCCA	CTACAAATAT	3420
TATAGAGCCG	AAAATTCACA	TCTAATATAT	GCAGACTACT	TTGAAATGAA	ATTAAAAAAA	3480
TTATTAAAGG	ATGACACAAA	AGTTTTTGAA	AAATCTACAT	TCAAATTTGT	AGAAGGATAT	3540
AAAATATACC	TGACAGAATC	TAAAGAATCT	GGAATTAAAC	AAATGGACAA	TGTCATAAAA	3600
TATTTTGAGT	TTATTGAAAT	TAAAAGTATT	GCTTTATATT	TTCAAAAACG	ATTAAATGAG	3660
CTGATAGATT	AAATAGCATT	TTCTCTGTTG	AGATATTGTT	TTTAAAAATAT	TGTAATAAAT	3720
GATTGATGCT	ATGTGGAAAT	ACAAAAAAT	GTTTTTGATA	CGAAGTTGAC	CTGTATTTTT	3780
TATACTAATC	ATTTTCGTAT	TTTTTGATTT	AAACGATATA	AGTTTGTGTT	AAACTTACAA	3840
GGAATAAAGA	CATTAAAAAA	TAACAGTATA	TCTATTGTTT	TTATATATTT	TACGAATTCT	3900
GCATAAATCT	CTTTCTAGTA	ATGTGTTGTA	ACTCTGCTAT	AATAGATTTA	TTCTTTTTTG	3960
TGTTTACACA	ATTTATTTTA	TACTACCAAA	AAAGGTCAGG	ATTTGTTCCT	TGACCTTTGA	4020
CAACTTTACC	GATTCTTTAG	TTCTACATAG	CGCTTGATCC	AAATGTTTAC	ATAGGCTTCT	4080
GAGAAAGGAC	CACGTCCATT	GTAAATCCAA	TCAACAAGAA	TTTTGACATG	TTCTTTTAAA	4140
ATATAGTCCA	AGTCATCAGA	ATAATTCATT	TTGCGTTTGT	GACGCTCGTA	CTCTTCAACG	4200
TCCAAGAGAC	GTTTTTCCCC	ATCTGTAAAA	ATTTTAACAT	CCAAATCGTA	ATCAATATAC	4260

631

TTCAGTGCCTT CTTCATCCAG ATAGTAGGGG CTAGCCATAT TGCAATAGTA AGAAGTTCCA	4320
TTATCACGAA TCATGGCAAT GATATTAAAC CAATATTTCT TGTGAAAGTA AACAAATAGCC	4380
GGTTCTCGAG TGACCCAACG ACGACCATCA CTTTCGGTAA CAAGTGTATG ATCGTTGACA	4440
CCAATAATGG CGTTTTCTGT TGTTTTTAGT ACCATGGTGT CCCGCCAAGT TCGGTGGAGA	4500
CTCCCATCAT GCTTATAACT TTGAATTGTA ATAAAGTCGC CTCTTTTGG AAGCTTCATA	4560
ACTAACCAAC TTTCTACAAT TTATAAGTTT ATCATTACT ATTGTACCAT AAAATTACCC	4620
AAAATCTGTG AATTTCACTT GGAAATATTA AAGATATTCT CTAAGAGCGC TTGCTATATC	4680
CGAAAAATCG TAGCCCTTTC GTGCTAAAAC TTGAGTTAAA CGCTGCTTCA GTTCGTATCC	4740
TTCATACTTT CGGGCATACT TAGTATATTG CTTATCAAGT TCCTTGAAGA TGAGTTCCTG	4800
AGTCGTTTCT TCATCAACTT GACTATCCAA TTCGTCAAAG GCAATTTTAG CATCAAAATA	4860
AGAGAAGCCC TTGTTAGTCA AGTCTGGAT AATCTTATCT TGCAGGGCAC GAGCTGGAAG	4920
TTTTCCCTCA TATTTTTTCA ATAGTTTATT GGCTACACGT TGAGCAACTT CCGAAAAATC	4980
AAAATCATTC AAGATTTCTT CTATAGTAGA TTTTGAAATT CCTTTTGTG CTAATTTCTG	5040
AGTCAGTACA TAAGGTCCCT TGTCTCCTGA AAGTTGATTG GCATTGATGA TAGCATAAGC	5100
GTACTGGCTA TCATTAATCC ACTTCTCTTC TTTAAGATTA GCAATGACTT GAGAAACGAT	5160
GTTTTCATTA ATATCATATT TTTTCAGATA TTCTCTGACC TCTTTTTCAG TACGTGCTTT	5220
AAAGGATAAG TGGTAGAGGG CCAGATTCTT ACCATAAGAA AATTGAGCAA AGTCTTGAAT	5280
CTCTTTCAAT TCCTCTTCGC TTATCACCTT ATCTCTCGAT AACATAAAAC GAACAATTGT	5340
GTCTTCGGTG ATATAGCATT TGTCG	5365

(2) INFORMATION FOR SEQ ID NO: 78:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3636 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TTTCCAGAAA GAAGTTGAGT AAAGTCTTTA TCAAAGAGAA TGACTTCGGT ATTGGAAGTG	60
ACATTAGGTT TTATTTCTAC TTTACTAGCG TCCGCCCTAG CATTTTCTAA ATCTTTAATC	120
TCTTCTGTTG CCTATTTAT AGCCAGCTGA ATAACTGCTT GAGGATTTTC ACTCAGTCCA	180
TGAAGCTTAT CGTCCACCGA AGTATAAAGA CTCGAATGCA TGACTTGTA AATAATCAGA	240

612

GTCATTGTAG AAAAAATCAG GGTGAAGACA CCGAAGTTGC GGATAAAATA ACTAAAGTCA	300
TCCGCATACC ATGTTTTTTT AAGTTTACTG AACATCTTTT AAAAGATACC CAACACTACG	360
CAAAGTTTGC AAATTCTCTG CAAAAGTGGT TCCCTTTAAT TTCTTACGGA CTTTGTAAAC	420
ATAGACTTCG ACAACCGAAA TCGTTGTATC ACTATCAAAT CCCCATAGAC GGTCAAAAAT	480
CTGCGTCTTA GGCAAAATCA CATTTTGATT TTGAAGGAAA TAAACTAGTA AATCGAACTC	540
TTTCCCCAGC AATTGACAG GAGTATCTTC AACTTTAACG GTATTGGTTG ATAAATTAAC	600
CACGATATTC CCATAAGTCA AGGTGTTTTT ATTAACTTC CCTGAACGTT TGAGAAGGGC	660
CTGAATCCGC ATTTTAAGTT CTTCTAGGTA GAAAGGTTTG GTCAGATAAT CATCCGCTCC	720
CAGTTCAAAT CCATGTCCCT TGTCAATCAA ACTTTCCTTG GCAGTCATAA TCAGAACTGG	780
TGTCGTAATT CCCTTTTCAC GCAATTCTTT TAAGACTTGG AAACCATTTT TTTCTGGCAA	840
CATCAAATCC AGCAAAATCA AGTCATAGAC ACCACTCTCA GCTTCGTAGA GACCTTCTTC	900
TCCATCAAAT ACCTGCATAA CATCCGCAAA ATCGTCTAAA AAGTCAAATA CTGAATTTGA	960
CAGACCTAGG TCATCCTCAA CCAATAAGAT TTTTATCATG AGAAACTCCT CTTTATTAAA	1020
ACTATTATAC CAAATTGCCC TTAACAAAAA CTCAACTCTC TGCATTTTAC ATGAGATAGC	1080
TGAGTTTCT TTTTATTTA GGCTTATTTA TGCATTTCCG TATTGAAGAA CAACTGCTTC	1140
GACTGCAGCT TTTTCACGGC TAATCAAGTC AACACGCGCT GCAATTTCTT TGATTTCCAT	1200
ACCGATGTTA CGGCTAAGAG CAAGGTCAGA AAGTTGCGGT TCAAAGAACT CTTTGTATTC	1260
CGCCAAGCGT TGCTGAGTCT TAAATACATG AGCAGGAAGG ATAACAAAGC TATCAAAGCT	1320
CATATCTCCT CCAAGGGCTG CCTTAATCCA AGCCCAGTTT TCACGCGCCC AAGACCAAGC	1380
TGTTTTCTGA GTTGCTTGAT GAGCTAGGAA TTGGTAATAC CAAGCAGACA AGTCTGTGG	1440
TTTGACCACA AATTTGTCCT TCCAAGAAGT AATCAGGTTT TGGATATTAT CCGCATCTGT	1500
ACTGTATGCA AGAGCTGCTG CCAACTGGCG TTTAAAGACA GCATCTGTTG CGTGAGTATA	1560
AGTATCAAGA TAAAGTGCTA ACAAGTCTTT AGTCTCATGA TGTTCATCT CATTAATCAG	1620
AACTTGTGAG CGAATAGCTG CTGGGAGTCC TGCAAGATTC TCCTTGTGTG TTGCGAAGAT	1680
TTGGCTAGCG ACTTGACTAG CTTCTGCATC ATTTGAGCGA ATCATCATCG AAACAGCCAG	1740
CTGACGAACC AATTCATCCT CATCTGATTC TCCGTCTTTA GCTTCAAAAC CAAGACGGTC	1800
ATAGTTATGA CGAGCCAATT TAGCAACCAG TCCTTTGAAG GCTGTTTCAG CATCCGTTCC	1860
TTTATCAATA AAGCGCTCAA GGGCTGAAAT CACTTGAGAA ACAGCTGAAA CCACCAGATA	1920
AGACTCTTCC TTAGCAAGTT TATCAAGAAC TGGAAGCAAG TCTGCATAAG AAATGTGCCC	1980
TGCCTCAGCC AACAAACGAC GTTCTTGAAC AATTGTCAGT TTGCTTGTGT TATCAAGTGT	2040

633

CTCTAGCTCA GCAAGAACAG CTGCTAACAA GTCTCCTTGA TAGTCGGTAA TATAGTGGGC	2100
AGTATTTTCA GTGTTGAGAC GAAGAGCTCC TTCATTTTCA GCAAGAAGAG CTGCGTAGCC	2160
AGGGATTTTCG ATACTTTTCAG TTTCGAGTGT ATCAGGCAAG CCTTTCCAGT TGCTATTGAG	2220
GGGCACCACC CAGAGACGGT TCTTGTCTTC GTTCTCACCG ATGAAGAATT GTTTTTGTGA	2280
AATCTTCAAG ACATCATTTT CAACTTTAAC AGTAAGAACT GGGTAACCAG GCTGTTCCAA	2340
CCAAGAATCC ATGAAGGCTG CGACATCACG TCCTGACGCT TGACCAAGGG CATCCCAAAG	2400
GTCACTACCA ATGGTGTGTC TGTATTGGTG TTTTCAAAG TAGGCGTGCA AACCTTTAGC	2460
AAAATCAGCA TCTCCTAGCC AACGGCGAAG CATGTGCATG AGACGGCTTC CTTTGGCATA	2520
GACGATAGCG CCGTCAAAGA GTGTATTGAT TTCATCTGGA TGTTTAACTT CGACGTGGAC	2580
AGACTGAACG CCATCAGTAG CGTCACGTTT AAGAGCAAGA GGTACTCCAC CTGTTTGGAA	2640
ATCTTCAAAG ATATTCCAGC TTGGTTCGAT GGTATCCACA CAGACGTATT CCATCATATT	2700
AGCGAAACTT TCATTGAGCC AAAGGTCATC CCACCATTTC ATAGTCACGA GGTTCCEAAA	2760
CCATTGGTGA GCCAATTCAT GGGCCACAAC AAGGGCAACT TGTGACGGC TAGCAAATGT	2820
AGAGTTCTCA TCGACAACCA AGTAAACTTC ACGGTAGGTC ACAAGACCCC AGTTTTCCAT	2880
AGCACCAGCT GAGAAGTCAG GAAGGGCGAT GTGGAGAGAT TGAGGAATTG GGTACTTAAC	2940
TCCATAGTAA TCTTCGTAAA ACTCGATAGA GCGAACAGCG ATATCCAGTG AGAAATCAAG	3000
ATTTGAAAGT GGATGTGCTT TGTTGAGTA GACACCTACC AGGTACCAT TTTTAGTTTT	3060
AGCGGTCACC CCTTGCAAAAT CACCAGCAAC AAAGGCCAAC AAGTAAGAAG ACATGCGAGG	3120
TGTTGTCTCA AACTTCCAGA TACCTGTTTC CTTACGGTTT TCAACATCGA TTTCTGGCAT	3180
GTTTGACAAG GCCAATTCAC CTTCTGCTTG GTCAAAGCGA AGAGAGAGGT CAAAAGTTGC	3240
TTTGGCTTCA GGCTCATCCA CACATGGGAA AGCTTCGCGC GCAAAATGGC TCTCGAACTG	3300
AGTAGACAAG ACCTCCTTCT TGAATCCATC AACTGTATAA TAAGAAGGCT AAATCCCTGT	3360
CATGTTGTCT GTAATTTTAC CAGAAAAGGC AAGAACCAAT TCAACTTGAC CAGCCTCAGC	3420
CAATTCGATA TGAAGGGCTT CATTGTCATG GTCAACTGTA AATGGACGAG CTTGACCTGC	3480
AACTTCTACA GAGGTGATTT CCAAATCTTT TTGGTGGAGG GAGATGCGGT CACTCTGTGC	3540
TTGACCAGTG ATGGTCACTT TCCCAGAAAA AGTCTTGGTC TCACGACTCA AATCTAAAAA	3600
TAAATCATAA TGTTCAAGAA CAAATTGCTT AATGGG	3636

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5066 base pairs

634

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ATAGCGTGTATAAATCGATT	TTAGAGGTAC	CATAAGCCAC	CTCCTACAAA	TAGAAACCGA	60	
TATAAATCAA	TGCCTTCCAC	CCTTAGACTT	CCCTAGTTCC	TGTCTCAAGC	GAAACATTTT	120
TTTGAAACAG	GAATAAGTTA	ACCAATTCAT	ACCAATAGCT	AGCAGAATAA	AAAGAAACCA	180
AATGCCCCAT	AACCTTGATAT	CTGTCACATT	TCTCAAGACG	GTATTGAAAA	ACAGAACTGA	240
AACAACTGTC	CAAGCAAGGC	TAAAAAGAGA	ATAGAAGGGG	ATGTAAAACC	AGTAAAAATA	300
ATAAAAAATT	GGAAAAAAT	TACTATTCT	GTTGGCCTTT	TCAATCCAGT	TATCAAAATA	360
AAAGTACGGT	GCTAAAAGTA	AGAATTTAAA	CAAATGTTCC	ATCACCGACA	TCCCCCCTTC	420
TTTTGATAGC	GTTTTCTATT	ATTTTATTAT	ATCAAAAAAA	TCCGGAACGT	TCATTCCAGA	480
TTCTACTTTT	TTATTTGCGT	TTTCTTGCGA	TGAGATGAAT	CGGTGTTCCC	TCAAAAACAA	540
AGGCCTTGCG	GATTTGATTT	TCCAAGAAAC	GCAGGTAAGA	AAAGTGCATG	AGTTCTTCTT	600
CATTGACAAA	GATGACAAAG	GTTGGTGGTT	TGGTTGCCAC	TTGGGTGCGA	TAGAAAATCT	660
TGAGACGTTT	TCCTTTGTCT	GTCGGTGTG	GGTTGATGGC	AATGGCATCC	ATGATGACAT	720
CGTTCAAGAC	AGCTGATGGA	ATACGTGTAT	TTTGACTTTT	GCTGATTTGC	TTAATCATCT	780
CAGGAAGTTT	GTGGAGACGT	TGCTTGTTA	AAGCTGATAC	AAAGATAATC	GGTGCGTAAG	840
GCAGGTATTG	GAAGTGTCTA	CGGATATCTT	CTTCCCAGTT	TTTCATAGTG	TGGTTATCTT	900
TTTCAAGCGT	ATCCCACTTG	TTGACCACGA	TAATCATCCC	TTTACCAGCT	TCATGGGCAA	960
ATCCTGCGAT	ACGCTTGTCG	TACTCAGGAA	TGCCTTCTTC	CGCATTGATG	ACCATCAAGA	1020
CCACATCTGA	ACGGTCAATA	GCACGCATGG	CACGCATAAC	AGAGTATTTC	TCAGTATTTT	1080
CATAAACCTT	ACCAGACTTA	CGCATACCAG	CCGTATCAAT	CATGGTAAAC	TCTTGACCAT	1140
CTGTATCTGT	AAAGTGGGTA	TCAATGGCAT	CACGAGTTGT	TCCAGCAACA	GGACTAGCAA	1200
TAACACGGTC	TTCTCCCAAG	ATAGCATTGA	TCAAGCTTGA	TTTTCCAACG	TTAGGACGAC	1260
CAATCAAGCT	AAACTTAATG	ACATCTGGAT	TTTCTTCCTC	ATATTCATTT	GGAAGATTTT	1320
CTACGATCGC	ATCTAGCACA	TCCCCTGTAC	CGATTCCATG	GACAGATGAG	ATAGGCAATG	1380
GTTACCCCAA	ACCGAGAGCA	TAGAAATCAT	ATATATCATT	TCTCATCTCA	GGGTTGTCCA	1440
CCTTGTGAC	TGCGAGGATA	ACTGGTTTGT	GGGTCTTATA	AAGCTTACGA	GCTACGTATT	1500
CGTCTGCATC	AGTAATTCCT	TCCTTACCAG	ACACGACAAA	AACGATAACA	TCTGCTTCTT	1560



635

CCATGGCAAT TTCTGCCTGG TGCTTGATTT GTTCCATGAA AGGAGCATCG ACATCATCAA	1620
TTCTCCTGT ATCAATCATG CTAAGAAGAAC GATTGAGCCA CTCACCCGTT GCATAAATAC	1680
GGTCACGTGT CACTCCTTCG ACATCTTCTA CAATGGAGAT TCGCTCACCA GCGATCCGAT	1740
TAAATAGGGT TGATTTCCCA ACATGGGGAC GTCCTACAAT GGCAATAGTT GGTAGGGCCA	1800
TAATTTCTCA CTTTCTACAA TAATTTCTTC TGTTCAGAT TTTTCTAGT TGAGCTTGGT	1860
TCAGCTTGAC CAACTGTTT TGCTAGGCGC TGAATCCAGC TTGTGGTCCG ACGCGCCCCA	1920
GCATAGTCAG CCTGAACACG GTCATAAGCT TGGATTGCCT CAGTTGACTG TTCTTGGTAT	1980
TCTTCCTCAA AGACAACATT CTCTAGTGGC AGTCTCGGTT TCATATCATG ATGTTGATTT	2040
GGCACACCCA GTGCCATCCC AAAGACAGAA TAGGTGTAGT CAGGTAGGTT AAAGAGCTCT	2100
GCCACTTCTT CAGACTTGTA TCGAACCAAA CCGATAATCA CACCACCATA GCCCAAGCTT	2160
TCAGCTGCCA ACAAGGCGTT TTGTCCAGCA AGAGCTGCAT CGACCGAACT AATCAAGAGA	2220
CCTTCCACAC CTTGGGGTTG GAAGGTGTCG GTATGAAGTC GGGCTCCCTT TTCTGCTCGG	2280
TTCAAATCTC CGACAAAGAG AAGGAAAACA GCAGACTGGC GAATGGCTTC TTGAGGTACC	2340
AATTCATACA AGGCATCTTT CTCTCTTGA CTTCTGTACCA CAATCACAGA GTAGGATTGG	2400
AAATTTCTCC AAGATGATGC CATCTGGGCT GCTGTCAAAA TCTCATTTAA GTCTACTTGG	2460
GGAATTTCTT GCTCTTTAAA CCTGCGCACT GAAGTATGAG CCTTCATCAA TTTAATGGTT	2520
TCTGTATCG ACGGTTTACT CTTCTAAAC GAGTCTCCTC AGCCAAATAA CGGATGCGTT	2580
CCATGACCCG TCTGGCTTCC CAGGTTTCGT CATTTCCATG TTCACTTTC GCAAAATGCT	2640
TCTCCAAATC TTCAAAGTTG AAGTTGGATG TGAAAAAGGT CGGTAAATTT TCCTGCATCC	2700
GATATTGGAG AATGACCTGC AGGATTTTCGT CACGCACCCA AACGGTTGAT TGCTCGGCGC	2760
CAATATCATC TAAAATCAGG ACCTCAGACA GCTTAATCTC ATCCACCAAG GTCTTAACAT	2820
TGCCATCACT GATAGCATTT TTGACATCAA TGACAAAGCT AGGATAGTGG AGGAGAGTTG	2880
ATGAAACACC ACGTTTTTCT GATAAATCAT GAGCTAAGGC CGCCACCATG AAACCTTTTAC	2940
CCACACCAAA GTCTCCATAT AAGTAAAGAC CTTTTCGAAT AGCTGGATAT TGCTCCACGA	3000
AGGCTAGTAG CTTTTCAAAA ACTGGTAAGC GCCCAAAATC ATCCAAGTCA ACTTGAGCCA	3060
AACTAGCTTT CTTGAGACTG GCTGGTAGAT TGATTAACCT GAGACGGTTC TTAATAGCCG	3120
CTTCTTTTTC AGCCGCGATT AGCTCAGGAG TTTCTTCATA TGAAACATCT GCATAACCAT	3180
GATTCTTAAC CAAAATCCGC TTGTAGCCTT TGGCAATATA ATCCGTATCC CCACGGAGAA	3240
ACTGTGCACG CTCGGTGATG TACTGATTAA ACTGGAGAT ACTGCGATTT AATTCCTTTG	3300

636

GAGTTAAGGA TTCTTGCTGG ATAAAGGCCG CAACATCAGG GTCCTTCATG ATTTTCTGGA	3360
CCAAATCTTG ATAATAAAAA CGGCTGGGTT GACGTTTGAG TACGTCTCCG ACACTTTCCA	3420
TCTAATCTCC TCCTTTTTCT AATCGAGCTA ATAGTTCTTG CTTCTTACGT TCTAGTTCCA	3480
GACGAGTTTC CTCGCTGGTT TCATTCTTAT ATTCAGGATT ACTCCATTTA GGAACATTGG	3540
TTTTTTCTGG GGCAGTCTGA TTCTGTTTTT GTGTTTTTGC TTTCTGCCCT CGATCACGAA	3600
TTTCGTA AAC GGCCTCTTCT GCCGAATGAA TCTTTTGATA GGCATAGTCA TTGGCTACCT	3660
TCATGGCATA TTTCTCATTG ATATTGCGG AATCCACCTT ATTAAAGGTC AATAAGAGAA	3720
TAATATTGAT GACTTCGTCC AGTAAGCCCA AGCCAGCCAT CTGTTGCAAG AGTTCTCTTT	3780
CTGTTTGGGT AATGGTTCCC TTGCGTGTTT GCTTGATTTC TGCTAAGAAC TGCAGGGCAG	3840
TTTTACTTTT AGCTTCTTTG ATAATGGTCG CTTCCTTAAG ACTAAAGTCA GAGGAACTG	3900
GTTTTTGAGC AATTTTTTCA CGCATGCGTT TGGTTGAAAT AACCTGGGAA ACAGCTGTTG	3960
ACTTGGCCAA TTGATAGGTT TCAAACCAAG TCCATTTCTT CTCCTCGGCA ATAGCAAAGA	4020
GGTTTAAGAC ATCGGACTGC TCATCCGCAA AACGAAGTCC ATCTCGAGCC ATCAGCTGGC	4080
GAAATGTTT CAAGTCAAAA TCATTGGCCA CTTTCTTCTT GAGACCAAGG TCTTCTTGAC	4140
TGCCTAGTTC TGCCAATTCT GGAAAGACTT GATTGAGTGA GACAGGTATT TCTTCACCAT	4200
CAGCACTTTC AACTTTCAAA TCCTCCACAG CTACATCGCC AATCTTTTTT TCTAAGAGTC	4260
TGCGATAAAC AGGATGCCCC AAGAAGTCTT GACTAGATAG AGGAGCATGG AGGGCTAGCT	4320
GATAAACATC ACCCTTTTGA TAGAGGGTCA AGAGATTAAA AGCAGATAAG ATTTTCAATG	4380
ATTTTATCAG TCTATCCATC CCAAAGTTGA GATGGTTGAG AATGCTTGAA AAAAGATATT	4440
CCTTTCTACC ATTATCCCAA AACTGATTG TATAAGATA AAGGCTCAGT GCCTCCTGAC	4500
CGATAATCGG GAGGTAGCAC TGTACCAGAG ATGAGGTATC TTGCGACACC CGATTATTCT	4560
TTAGATAAGA AAAACGGTCA ATTGGCTTCA TTTATCTTTC CTTTTTCTTT TTAGAGGACT	4620
GGGTGATTG TTGGAGCAAG CTCTCTAACT CACTGACATC CTTAAACTA CGATAGACAC	4680
TAGCAAAACG TACATAGGTA ATCTCGTCCA ATTCAGCCAA CTCCTCCATG ACGAGTGAAC	4740
CAATGTCCTC ACTTTGAATT TCATTTTCAT TTCGACCACG GAGTTTCTGT TCGATACGAT	4800
TGACTACCAT GTTGATTTC TCACTTGACA CAGGACGTTT CTGGGCTGAG CGGATAATCC	4860
CATTAAAGAT TTTATCTCTG GAGAATTGTT CCCGTGTGCC ATCTTTTTTA ACAACCACTA	4920
AGGTTCTTTC TTCTACTCGT TCGTAGGTTG TAAACGGTG TTGGCATTG TCGCACTCAC	4980
GTCTTCTACG AATGGTGTTT CCTTCTTCTG CTTGGCGACT ATCGATAACA CTTGACTTGG	5040
TAGCCCCACA TTTTGGACAG GGTACC	5066

637

## (2) INFORMATION FOR SEQ ID NO: 80:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9607 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

CACCTGAAGT ATTTGAAACA GCTATGGAAC ACATCATGCC TGTACTTGAA GTACGTGCAC	60
GTCGTGTTGG TGGTTCTAAC TACCAAGTCC CAGTTGAAAGT TCGTCCAGAA CGTCGTACAA	120
CACCTGGACT TCGTTGGTTG GTAACAATCG CTCGCTCTTCG TGGTGAACAC ACAATGCAAG	180
ACCGTCTTGC AAAAGAAATC TTGGATGCTG CTAACAACAC TGGTGCAGCA GTTAAGAAAC	240
GTGAAGATAC TCACCGTATG GCTGAAGCTA ACCCTGCATT CGCACACTTC CGTTGGTAAG	300
ATAGGATGCG AAAGCGTTAA GAAAGTCCCA GAGAAAATAG GGAATCGAAG CAGGTTGCGG	360
TTGCAACCAA TGAGATTCAT CTTTTTCTCC AGACTTTTAG CTTGAGCTCA ACTAAATCAT	420
GATGCTAGGA ACGGTAAGGA TGCAAGGTAA AAATAGGAAA CTGACGCAGT ATTCGACGAA	480
TACAAGGAGT TTTATCTTTT TCACGCAGCA TCCCGTTCCA GCTCACATCG CCTAACTAAC	540
TTTAGCCCGG GTTCAAATTA GCTAAATCGA TTAGTATTAG CTATAACTCA GCTTACCATC	600
TCGTAAGTTG AAACCAACAA TAGCATGAAA ACATTGAGAA CCGGTAGGTC CTGCCTATCC	660
GTTTTTATTA AAATCGTGTT ATAATAGAAT AGAAATCAAA AATAAATAGG AGAAACAAAC	720
CTCATGGCAC GCGAATTTTC ACTTGAAAAA ACTCGTAATA TCGGTATCAT GGCTCACGTC	780
GATGCCCGTA AAACAACAAC TACTGAGCGT ATTCTTTACT ACACTGGTAA AATCCACAAA	840
ATCGGTGAAA CTCACGAAGG TCGGTCACAA ATGGACTGGA TGGAGCAAGA GCAAGAACGT	900
GGTATCACGA TCACATCTGC TGCGACGACA GCTCAATGGA ACAACCACCG CGTAAACATC	960
ATCGACACAC CAGGACACGT GGACTTCACA ATCGAAGTAC AACGTTCTCT TCGTGTATTG	1020
GATGGTGCGG TTACCGTTCT TGAATCACA TCAGGTGTTG AGCCTCAAAC TGAAACAGTT	1080
TGGCGTCAAG CAACTGAGTA CGGAGTTCCA CGTATCGTAT TTGCCAACA AATGGACAAA	1140
ATCGGTGCTG ACTTCCTTTA CTCTGTAAGC ACACTTCACG ATCGTCTTCA AGCAAATGCA	1200
CACCCAATCC AATTGCCAAT CGGTTCTGAA GATGACTTCC GTGGTATCAT TGACTTGATC	1260
AAGATGAAAG CTGAAATCTA TACTAACGAC CTTGGTACGG ATATCCTTGA AGAAGACATC	1320
CCAGCTGAAT ACCTTGACCA AGCTCAAGAA TACCCTGAAA AATTGATTGA AGCAGTTGCT	1380

638

GAAACTGACG	AAGAATTGAT	GATGAAATAC	CTCGAAGGTG	AAGAAATCAC	TAACGAAGAA	1440
TTGAAAGCTG	GTATCCGTAA	AGCGACTATC	AACGTTGAAT	TCTTCCCAGT	ATTGTGTGGT	1500
TCAGCCTTCA	AAAACAAAGG	TGTTCAATTG	ATGCTTGATG	CGGTTATCGA	CTACCTTCCA	1560
AGCCCACTTG	ACATCCCAGC	AATCAAAGGT	ATTAACCCAG	ATACAGACGC	TGAAGAAATT	1620
CGTCCAGCAT	CTGACGAAGA	GCCATTGCA	GCTCTTGCC	TCAAGATCAT	GACTGACCCA	1680
TTCGTAGGTC	GTTTGACATT	CTTCCGTGTT	TACTCAGGTG	TTCTTCAATC	AGGTTTCATAC	1740
GTATTGAATA	CTTCTAAAGG	TAAACGTGAA	CGTATCGGAC	GTATCCTTCA	AATGCACGCT	1800
AACAGCCGTC	AAGAAATCGA	CACTGTTTAC	TCAGGTGATA	TCGCTGCTGC	CGTTGGTTTG	1860
AAAGATACTA	CAACTGGTGA	CTCATTGACA	GATGAAAAAG	CTAAAATCAT	CCTTGAGTCA	1920
ATCAACGTTT	CAGAACCAGT	TATCCAATTG	ATGGTTGAGC	CAAAATCTAA	AGCTGACCAA	1980
GACAAGATGG	GTATCGCCCT	TCAAAAATTG	GCTGAAGAAG	ATCCAACATT	CCGCGTTGAA	2040
ACAAACGTTG	AAACTGGTGA	AACAGTTATC	TCAGGTATGG	GTGAACTTCA	CCTTGACGTC	2100
CTTGTTGATC	GTATGCGTCG	TCAGTTCAAA	GTTGAACCGA	ACGTAGGTGC	TCCTCAAGTA	2160
TCTTACCGTG	AAACATTCCG	CGCTTCTACT	CAAGCACGTG	GATTCTTCAA	ACGTCAGTCT	2220
GGTGGTAAAG	GTCAATTCGG	TGATGTATGG	ATTGAATTTA	CTCCAAACGA	AGAAGGTAAA	2280
GGATTGCAAT	TCGAAAACGC	AATCGTCGGT	GGTGTGGTTC	CTCGTGAATT	TATCCCAGCG	2340
GTTGAAAAAG	GTTTGGTAGA	ATCTATGGCT	AACGGTGTTT	TTGCAGGTTA	CCCAATGGTT	2400
GACGTTAAAG	CTAAGCTTTA	TGATGGTTCA	TATCAGCATG	TCGACTCATC	TGAAACTGCC	2460
TTCAAGATTG	CGGCTTCACT	TTCCCTTAAA	GAAGCTGCTA	AATCAGCACA	ACCAGCTATC	2520
CTTGAACCAA	TGATGCTTGT	AACAATCACT	GTTCCAGAAG	AAAACCTTGG	TGATGTTATG	2580
GGTCACGTAA	CTGCTCCTCG	TGGACGTGTA	GATGGTATGG	AAGCACACGG	TAACAGCCAA	2640
ATCGTTCGTG	CTTACGTTCC	ACTTGCTGAA	ATGTTGCGTT	ACGCAACAGT	TCTTCGTTCT	2700
GCATCTCAAG	GACGTGGTAC	ATTCATGATG	GTATTTGACC	ACTACGAAGA	TGTACCTAAG	2760
TCAGTACAAG	AAGAAATTAT	TAAGAAAAAT	AAAGGTGAAG	ACTAATCCGT	CCTCACTCTA	2820
GAAGGAAGTC	ACTTAGTGCC	TTCTTTTGT	CTTTAGAAAA	TACCTCTAAA	TATGGTAAAA	2880
TAGTAGAAGA	ATAATGTGAG	GAAAAATGAAT	GTCAAATAGT	TTTGAAATTT	TGATGAATCA	2940
ATTGGGGATG	CCTGCTGAAA	TGAGACAGGC	TCCTGCTTTA	GCACAGGCCA	ATATTGAGCG	3000
AGTTGTGGTT	CATAAAATTA	GTAAGGTATG	GGAGTTTCAT	TTTGTATTTT	CTAATATTTT	3060
ACCGATTGAA	ATCTTTTITAG	AATTAAAGAA	AGGTTTGAGC	GAAGAATTTT	CTAAGACAGG	3120
CAATAAAGCT	GTTTTTGAAA	TTAAGGCTCG	GTCTCAAGAA	TTTTCAAATC	AGCTCTTGCA	3180

639

GTCCTACTAT AGGGAGGCTT TCTCTGAAGG TCCATGTGCT AGTCAAGGTT TTAAGTCCCT	3240
TTATCAAAAT TTGCAAGTTC GTGCTGAGGG TAATCAGCTA TTTATTGAAG GATCTGAAGC	3300
GATTGATAAG GAACATTTTA AGAAGAATCA TCTTCCTAAT TTAGCCAAAC AACTTGAAAA	3360
GTTTGGTTTT CCAACTTTTA ACTGTCAAGT CGAGAAGAAT GATGTCCTGA CCCAAGAGCA	3420
GGAAGAGGCC TTTCATGCTG AAAATGAGCA GATTGTTCAA GCTGCCAATG AGGAAGCGCT	3480
CCGTGCTATG GAACAACCTGG AGCAGATGGC ACCTCCTCCA GCGGAAGAGA AACCAGCCTT	3540
TGATTTTCAA GCGAAAAAAG CTGCAGCTAA ACCCAAGCTG GATAAGGCGG AGATTACTCC	3600
TATGATCGAA GTGACGACAG AGGAAAATCG TCTGGTATTT GAAGGGGTTG TTTTGTATGT	3660
GGAGCAAAAA GTGACTAGAA CAGGTCGTGT TTTAATCAAC TTTAAATGA CGGACTATAC	3720
TTCAAGTTTT TCTATGCAAA AGTGGGTAA AAACGAGGAA GAGGCCCAGA AGTTTGACCT	3780
CATCAAGAAG AATTCCTGGC TCCGAGTTCG AGGGAATGTG GAGATGAATA ACTTCACACG	3840
CGATTTGACT ATGAACGTAC AGGATCTGCA GGAAGTTGTT CACTATGAGC GGAAGGATTT	3900
GATGCCAGAA GGTGAGCGTC GGGTTGAGTT TCATGCTCAT ACTAACATGT CGACTATGGA	3960
TGCTTTGCCA GAGGTCGAAG AGATTGTTGC AACAGCTGCT AAGTGGGGAC ACAAGGCGGT	4020
TGCTATCAGC GACCATGGGA ATGTCCAGTC CTTTCCACAT GGCTATAAGG CGGCTAAGAA	4080
AGCGGGAATC CAGCTGATCT ATGGGATGGA AGCCAATATC GTGGAGGACC GTGTCCCTAT	4140
CGTCTATAAC GAAGTGGAGA TGGACTTGTC AGAAGCAACC TACGTGGTCT TTGACGTGGA	4200
AACGACGGGA CTTTCAGCTA TCTATAATGA CTTGATTACG GTTCCGGCTT CTAAGATGTA	4260
CAAGGGGAAT GTTATTGCTG AATTTGATGA ATTTATCAAT CCTGGGCATC CCTTGTACGC	4320
CTTTACTACA GAGTTAACTG GAATTACAGA TGATCATGTC AAAAATGCCA AACCCTAGA	4380
ACAAGTTTTG CAAGAATTCC AAGAATTTTG CAAGGATACG GTCCTAGTTG CCCACAATGC	4440
TACCTTTGAC GTTGGCTTTA TGAATGCTAA TTATGAGCGG CATGATCTTC CAAAGATTAG	4500
TCAGCCAGTT ATTGATACGC TGGAGTTTGC TAGAAACCTC TATCCTGAGT ATAAACGCCA	4560
TGGTTTGGGG CCTTTGACCA AGCGTTTTGG TGTGGCCTTG GAACATCACC ACATGGCCAA	4620
CTACGATGCG GAAGCGACTG GTCGTCTGCT TTTCATCTTT ATCAAAGAGG TAGCAGAAAA	4680
ACATGGTGTG ACCGATTTAG CTAGACTCAA CATTGATCTA ATCAGTCCAG ATTCTTACAA	4740
AAAAGCTCGG ATCAAGCATG CGACCATCTA TGTCAAGAAT CAGGTAGGTC TAAAAAATAT	4800
CTTTAAGCTG GTTTCCTTGT CTAATACCAA GTATTTTGAA GGAGTGCCAC GGATTCGGAG	4860
AACGGTTCTA GATGCCCATC GAGAGGGCTT GATTTTAGGT TCAGCCTGTT CAGAGGTTGA	4920

640

AGTTTTTGAC	GTGGTCGTTT	CTCAAGGTGT	GGATGCGGCG	GTTGAGGTGG	CCAAGTATTA	4980
TGATTTTATC	GAGGTCATGC	CACCGGCTAT	CTATGCACCC	TTGATTGCCA	AAGAGCAGGT	5040
CAAGGATATG	GAGGAACTCC	AGACCATTAT	CAAGAGTTTG	ATAGAGGTTG	GAGACCGCCT	5100
TGGCAAGCCT	GTTCTGGCTA	CGGGAAATGT	TCACTATATC	GAACCGGAAG	AAGAGATTTA	5160
TCGTGAAATT	ATCGTCCGTA	GTTTGGGACA	GGGTGCGATG	ATTAATCGAA	CTATCGGTCA	5220
TGGTGAACAT	GCCCAACCAG	CACCACTTCC	AAAGGCTCAT	TTTCGAACGA	CTAATGAGAT	5280
GTTGGATGAA	TTTGCCTTTT	TGGGAGAGGA	ACTGGCTCGT	AAACTGGTTA	TTGAAAACAC	5340
CAATGCCTTG	GCAGAAATAT	TTGAATCCGT	TGAAGTCGTT	AAGGGTGACT	TGTATACGCC	5400
TTTCATCGAC	AAGGCTGAAG	AAACAGTTGC	TGAGTTGACC	TATAAGAAAG	CTTTTGAGAT	5460
TTATGGAAAT	CCGCTGCCAG	ATATTGTTGA	TTTGCGGATT	GAAAAAGAAT	TAACATCCAT	5520
ACTGGGGAAT	GGATTTGCTG	TGATTTATCT	GGCATCGCAG	ATGCTGGTGC	AACGTTCTAA	5580
TGAACGGGGT	TATTTGGTTG	GTTCTCGTGG	GTCTGTCGGA	TCTAGTTTCG	TTGCGACCAT	5640
GATTGGGATT	ACGAGAGTCA	ATCCTCTCTC	TCCTCACTAT	GTCTGTGGTC	AGTGTCAGTA	5700
CAGTGAGTTT	ATCACAGATG	GTTCTGACGG	TTCAGGATTT	GATATGCCCC	ATAAGGACTG	5760
TCCAAACTGT	GGTCACAAAC	TCAGTAAAAA	CGGACAGGAT	ATTCCGTTTG	AGACCTTCCT	5820
TGGTTTTGAT	GGGGATAAGG	TTCTTGATAT	TGACTTGAAC	TTCTCGGGAG	AAGATCAGCC	5880
TAGCGCCAC	TTGGATGTGC	GTGATATCTT	TGGTGAAGAA	TATGCCTTCC	GTGCGGGAAC	5940
GGTTGGTACG	GTAGCTGCCA	AGACTGCCTA	TGGATTTGTC	AAAGGTTACG	AGCGAGATTA	6000
TGGCAAGTTT	TATCGTGATG	CAGAAGTAGA	ACGCCTCGCT	CAAGGAGCGG	CGGGTGTCAA	6060
GCGGACAACA	GGCCAACACC	CGGGGGGAAT	CGTTGTTATT	CCGAACTACA	TGGATGTCTA	6120
CGATTTTACG	CCTGTCCAGT	ATCCAGCAGA	TGATGTCACG	GCTGAATGGC	AGACCACTCA	6180
CTTTAACTTC	CACGATATCG	ATGAGAACGT	CCTCAAATC	GATGTACTGG	GACATGATGA	6240
TCCGACTATG	ATTCGAAAAC	TTCAGGATTT	GTCTGGTATT	GACCCTAATA	AAATTCCTAT	6300
GGATGACGAA	GGCGTGATGG	CACCTTTTTT	TGGGACTGAT	GTGCTAGGGG	TAACACCTGA	6360
ACAAATTGGA	ACGCCTACGG	GTATGTTGGG	GATTCCAGAG	TTTGGAAACA	ATTCGTACG	6420
TGGAATGGTA	GACGAAACCC	ATCCGACAAC	CTTTGCGGAA	TTGCTTCAGC	TGTCTGGTCT	6480
GTCCACCGGT	ACTGATGTTT	GGTTGGGGAA	TGCTCAGGAT	CTGATTAAGC	AAGGAATAGC	6540
GGACCTATCG	ACTGTTATCG	GTTGTCGGGA	CGACATCATG	GTTTACCTCA	TGCATGCGGG	6600
TCTGGAACCT	AAGATGGCCT	TTACCATTAT	GGAACGGGTA	CGTAAGGGTT	TGTGGCTAAA	6660
GATTTTCAGAA	GAGGAGAGAA	ATGGCTATAT	CGAAGCAATG	AAGGCTAATA	AGGTGCCAGA	6720

641

GTGGTATATC GAATCCTGTG GGAAAATTAA GTACATGTTC CCTAAGGCCC ATGCGGCAGC	6780
CTACGTTATG ATGGCCTTGC GTGTAGCTTA CTTCAAGGTT CACCATCCTA TTTATTACTA	6840
CTGTGCTTAC TTCTCCATTC GTGCTAAGGC TTTTGATATC AAGACCATGG GTGCGGGCTT	6900
GGAGGTCATC AAGCGCAGAA TGGGAAGAAAT CTCTGAAAAA CGGAAGAACA ATGAAGCCTC	6960
TAATGTGGAA ATCGATCTCT ATACAACTCT TGAGATTGTC AATGAGATGT GGAACGAGG	7020
TTTCAAGTTT GGTAATTAG ATCTCTACTG TAGTCAGGCG ACAGAGTTCC TCATCGACGG	7080
GGATACCTT ATCCACCAT TTGTAGCAAT GGATGGTCTG GGAGAGAACG TTGCCAAGCA	7140
ACTGGTGCGG GCGCGTGAAG AGGGAGAATT CCTCTCTAAA ACAGAACTAC GCAAGCGTGG	7200
TGGACTCTCA TCAACCTTGG TTGAAAAGAT GGATGAGATG GGTATTCTTG GAAATATGCC	7260
AGAGGATAAC CAGTTGAGTT TGTGTGATGA GTTGTTTTAA AAAATTGCTT AATAATCTAT	7320
TAAAAGAGGC TAACGTATAT CCAATAGATT TACATTAGCT TTCTTTTGTG TAAAATAGT	7380
CTATGGAAAG AGGGTGAGAG TATGTCAAAG ATGAGTATAA GCATCCGTCT GGATAGTGAG	7440
GTAAAGGAGC AGGCCCAACA GGTGTTTAGT AATCTGGGAA TGGATATGAC AACAGCTATT	7500
AATATTTTCC TTCGTCAGGC AATTCAATAT CAGGGATTAC CTTTGTATGT TAGACTAGAC	7560
GAAAATCGGA AGTTGCTCCA AGCGTTAAGC GATTTAGACC AAAATCGTAA TATGAGCCAG	7620
TCTTTTGAAT CAGTCTCAGA TTTGATGGAG GACTTACGTG CTTAAGATTC GTTATCATAA	7680
ACAGTTTAAA AAAGATTTTA AGTTGGCTAT GAAGCGTGGT TTGAAGGCAG AATTATTAGA	7740
AGAAGTTTGT AATTTTCTGG TTCAAGAAAA AGAACATCCT GCCAGAAATC GTGATCATTC	7800
ATGACGGCA TCCAAGCATT TTCAAGGAGT TCGTAATGC CATACCCAGC CAGATTGGCT	7860
TTTGGTTTAT AAAGTAGACA AGTCGGAATT GATTTTAAAT TTGCTGAGGA CAGGCAGTCA	7920
CAGTGATTTA TTTTAATCTA TTTTAAGGGG GTTCTCATGA AACTAAGAAT ATTTGCGGAA	7980
GATAAGCCCG CTAAGAAGGT ATTTGAATAT CAATTAGAAC TTGCTGATCG TACAATTCTT	8040
CTATCGACAG CACTCTTGTC AGGTGCTATT GCTTTAGCAG GAATCTTTTC TGCTTTGAAA	8100
GAAAAATAAA AATAGAAAAG AGAAAACAGA ATGGTTTTAC CAAATTTTAA AGAAAATCTA	8160
GAAAAATATG CGAAATTGTT GGTGCGAAC GGAATTAACG TGCAACCTGG TCACACTTTG	8220
GCTCTCTCTA TTGATGTGGA GCAACGTGAA TTGGCACATC TAATCGTGAA AGAAGCTTAT	8280
GCCTTGGGTG CGCATGAGGT CATCGTTCAG TGGACAGATG ATGTGATTAA CCGTGAGAAA	8340
TTCTCCATG CCCCAGTGA GCGTTTGGAC AATGTGCCAG AATACAAGAT TGCTGAGATG	8400
AACTATCTCT TGGAGAATAA GGCTAGCCGT CTTGGAGTTC GTTCATCTGA TCCAGGTGCC	8460

642

TTGAACGGAG TGGACGCTGA CAAGCTTTCA GCTTCTGCTA AAGCTATGGG ACTTGCCATG	8520
AAGCCTATGC GTATCGCAAC TCAATCTAAC AAGGTTAGCT GGACTGTAGC AGCTGCAGCA	8580
GGACTTGAGT GGGCTAAGAA AGTCTTCCCA AATGCTGCGA GCGACGAAGA AGCAGTTGAT	8640
TTCCTTTGGG ACCAAATTTT CAAAACCTGC CGTGTCTACG AAGCAGATCC TGTTAAGGCT	8700
TGGGAGGAAC ATGCAGCCAT TCTCAAGAGC AAGGCCGATA TGCTTAATAA GGAGCAATTT	8760
TCAGCCCTTC ACTACACAGC GCCAGGAACA GATTTAACAC TTGGTTTGCC AAAGAACCAC	8820
GTTTGGGAAT CAGCTGGTGC TGTCAATGCA CAGGGCGAAG AATTCTTGCC AAATATGCCA	8880
ACAGAAGAGG TCTTCACAGC GCCTGACTTC CGTCGTGCAG ATGGTTATGT CACTTCTACA	8940
AAACCGCTTA GCTACAACGG AAATATCATT GAAGGCATTA AGGTGACCTT TAAGGATGGA	9000
CAAATCGTAG ATATCACTGC TGAGAAGGGT GATCAGGTTA TGAAAGACCT TGTCTTTGAA	9060
AATGCGGGTG CCGCTGCCTT GGGTGAATGT GCCTTGGTAC CAGATCCAAG TCCAATTTCT	9120
CAGTCAGGCA TTACCTTCTT TAACACCCTT TTCGATGAAA ATGCCCTCAA CCACTTGGCT	9180
ATCGGTGCAG CCTATGCGAC TAGCGTTGTT GATGGAGCGG AGATGAGCGA AGAGGAGCTT	9240
GAAGCTGCAG GGCTTAACCG TTCAGATGTT CACGTAGACT TTATGATTGG TTCTAACCAA	9300
ATGGATATCG ATGGTATTCG TGAGGATGGA ACGCGGGTAC CTCTTTTCCG TAATGGGAAT	9360
TGGGCAAATT AAGGAGATAA TATGTTAGGA AGTATGTTCC TTGGTCTCCT AGTGGGATTT	9420
TTAGCAGGTG CTATGACCAA TCGTGGAGAG CGAATGGGAT GTTTTGGAAT AATGTTTCTC	9480
GGTTGGATCG GAGCCTTTCT AGGTCACCTG CTCTTTGGAA CTTGGGGGCC AGTTTATCA	9540
GGAACAGCTA TTATCCCAGC GATTTTAGGA GCCATGATTG TTTTAGCTAT TTTTGGAGA	9600
CGAGGAA	9607

## (2) INFORMATION FOR SEQ ID NO: 81:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14231 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CTACAAGATA ATTCCAGCTA TAACATCCGC TATAATAGTA AGAGCGAGCT CTATGATAAG	60
GCTCATTAGT TTCACCTCCT CTCACGAACC CATAGGAACG TAATCGGTAA CCGATGACAA	120
AAATAGTATA CCACAATACA TTTAGATCAT CAAGGTCAC TAAATCTTGA AATATCAGAT	180
CTAAGAGAAA AATCTTTAAA ATCAGAAAAA CGCATAATAT CAGGTGTGCA AAAACTTGAT	240



643

ACTATGCGTT TTATTGTGGG AAGGTTTACT CCATTTTCTC CTGAAATTGA GTTTTGTGCC	300
AGCCTCTGTT TTTAGGGTTG CTAAGAAAAT AATGTCATGT GGTGAATATT TGTAAATCAG	360
TCAGCAGACA GAACGATACT CTTCGAAAAT CTCTTCACAT CATGTCAGCT TCGTCTTTCC	420
GTATATATGT GACTGACTTC ATCAGTTCTA TCTACAACCT CAAAACAGTG TTTGAGCTG	480
ACTTGATCAA TTTTCAAATC TGTACTTTGA GCAAGCTGAG ACTAGCTTCC TATTTGATTT	540
TCATTGAATA TCAGAAACCC ATTCTCCATC AAATAATTCT ACTGCGTCTA ATAATTTTTC	600
ATCTGGCAGC GTGTCTGAAA TAAAGGTTGT GTATTTGGAG AGGGGATTAA TTTTAAAAAA	660
TCCAGTCTTG TAAAATTTAG AACTATCAAT CAGTAAGATG GTTTCATGGG CTTTGTCAAT	720
AATATTCTTT TTTGAAATAG CTTGGCTGAG AGAAGCTTCA TAAACATATT GGTATCAAT	780
ACCTCTTGCT GAACAAAATG CTAATCGAT ATTAAAATGA TCTAATAAAG AATTTTCCTT	840
ATCATAGTTG ACCACGGAAC AGGATTGATG TTTGACCTCG CCAGATGTGA TAAACATTTT	900
GGAGCTATCT TTAACAGTTT CAGATAGGGT TTGTGCAGTA TGTAACCAT TTGTAAAAAT	960
AATCAAATTA TCAAGTTCAG AAAGATAGGG ACAGAGTTCC TAGACAGTAG TACTAGAATC	1020
TAGATAGATA CACATACCAG ACCGAATAAA GTCTTTAGCG AGACTAGCGA TTAGTCTTTT	1080
TTGCCTAGTA CTTTCTCCTT CACGTATTTG ATGAGAAAGT TCAATTGTGT TCATAGAGGA	1140
CAGGGTCACG TATCCGTGCT TTCTTTTGAT AAGACCTTGA TTTTCTAAGA AAATTAAATC	1200
ACGACGTAAG GTAAGTTGTC TGGAGAAAGT GATTTCTGCC AGCTCTTTTA CGGCAATTCT	1260
TTTTTCTTTT TTGATAATTT CAATCAATTC AAGTACACGT TCATCTTTTA TCATAAGCTC	1320
CTCCTAATTT ATCATTTCOA CTATATTATA GCACAAATTG GAGGAATTTG AATTATTTTT	1380
ATGAATATTG GGTTAACATT TGAACATTAT TCAAGTAAGC GTTCACATAT TGAAAAATA	1440
AAACGTGGGG ATTATAATAA AGTTAATCMA GGACGAAGAG AGAAGAAAAA TGGAAGCGGT	1500
TTTAGCAATA GATTTAGGTG CGACTTCTGG AAGAGCAATC GTTGGTTACC TTTCTGAAAA	1560
TAAACTAGTA ATGGAAGAAA TAAATCGCTT TTCTAATCTA CCTATTAGAG TAAAAGGGCA	1620
TTTATCTTGG GATATTGACT TTCTACTAGC TAAAATTCTT GAAAGTATCC GCTTGGCTAA	1680
TACTAGTTAC AAGATTTTAT CTATCGGTAT TGACACATGG GGAGTTGATT TTGGACTGAT	1740
TGATAATGAA GGTAAGCTGT TATTACAACC TGTTCAATAT CGTGATGAAA GAACAAAGGG	1800
AGTGTTAAAG GAAATATCTG AAATGACTGA ATTAGAAAAA CTGTATTCAG AGACAGGAAA	1860
TCAGATTATG GAGATAAATA CCTTGTTTCA ACTCTTTAAG GCACGTCAAG AATCTCCTGA	1920
CTCTTTCTAT AAGACCAATA AGATTCTTTT AATGCCAGAT TTGTTTAATT ATCTCTTGAC	1980

544

AGGTAAGTTT GCTACAGAAA AAAGCATTGC TTCAACAACCT CAATTATTTG ATCCTAGGAG	2040
TCAAAATTGG AATCAGAATA TCTTAAAACT ATTTGAATTG GATTCACTTT TACTTCCTGA	2100
AATTGTTTCA GAGGGAAATG TTCTTGAAG GATAAAAGAG GAGTATGGTT TAGGCGATAT	2160
TCCTGTGTG AATGTTTGTA GTCATGATAC AGCAAGCGCG ATTGTCTCAG TACCTAAGAC	2220
AGAAGGTAGT TTATTTATTT CATCAGGTAC TTGGTCTTTG GTTGGAGTGG AACTTACTTC	2280
ACCGATTCTT ACTACCGAAT CCTTCAGTTA TGGATTTACA AATGAAGTCG GTAAAGATGG	2340
AGTGATTACA TTTCTGAAGA ATTGTACAGG GTTGTGGATC ATAGAGGAAC TAAGACGTTT	2400
ATTTGAACGA AGAGGGAAAG CCTATTCTTT TGATGATATT AGGACAATGG TGGAGAAAGA	2460
AAAAGAAAAT CTTCCTCTGA TTGATACTGA ATCAACTGAA TTTGCAACAG AATCTGATAT	2520
GCACAAGACT TTGACAGAAT ATCTAGCTTA TCATCATGAA ACTAGAGAGT GGACAGATGG	2580
ACAACTATTT AAGATTGTTT ATGAAAGCCT AGCTGAAACG TATAGGAAAG CGATAGAGTT	2640
ACTAGAAGAA CTAACCTATA AGGTTTATAA GAGGATATAT GTGATTGGAG GAGGTGCTAG	2700
AGCCAGTTAC TTAAACCAAA TGATTGCTGA TAGAACTGGT AAAGAGGTTT TTACAGGTTT	2760
GACTGAGGCT ACAGCTGTGG GGAATATTGT TGTGCAGCTC ATAGCTATGG GACAATTAAA	2820
AGGGATGGAA GAGGCTCACC ATGTTATTGA GGAGTTTCTA CAATTAGAGA GTTATTACTC	2880
CCAAAAGAAT TAAAAAGATT GAGAGTTTGT AAATTTGCCT CCCTCCCCCT TCTTAGCTTT	2940
TGTGCAGGAA GGGGGGATAA TTGGTGAATT GAAAAATATT TAGTGTTTTG ATATGAGGAG	3000
GACAAGGATG TCAGATGTAA AACAAGAATT AATTAAATAT GGTAAGAAGC TAGTAGAAAC	3060
AGATTTGACG AAAGGAACAG GTGGGAATCT CAGCGTTTTC GATCGTGAAA AACAATTGAT	3120
GGCAATTACC CCGTCGGGTA TTGATTTCTT TGAATCAAA GAATCCGATA TTGTAGTGAT	3180
GGATATTAAT GGAAATGTTG TAGAGGGAGA ACGCTTGCCA TCTAGCGAAT GGTATATGCA	3240
TTTGATTCAA TATCAAACTC GTGATGATAT CGATGCAATT ATCCATGCTC ATACAACTTA	3300
TGCAACAGTA TTAGCTTGTC TCAGAGAACC ACTTCCAGCG AGTCATTATA TGATTGCAGT	3360
GGCAGGGAAA GATGTTCCGG TAGCTGAGTA TGCAACATAT GGCACGAAAG AATTGGCTGT	3420
GAATGCAGCT AAAGCAATGG AAGGTCGTAG AGCAGTTTTA CTAGCGAATC ATGGAATTTT	3480
AGCAGGTGCA CAAAATTTAT TGAATGCATT TAATATTGTT GAAGAAGTTG AATATTGTGC	3540
AAAAATTTAT TGTTTAGCTA AGAATTTTGG AGAGCCAGTA GTTCTTCCTG ATGAGGAGAT	3600
GGAATTGATG GCAGAAAAAT TTAACATA CGGTCAGAGA AAATAGGGAG GATATTTAATC	3660
TTAAACATA TACCGAAAAA TATTTCTCCA GATTTATTGA AGACTTTAAT GGAAATGGGA	3720
CATGGAGATG AAATAGTATT AGCTGACGCG AATTATCTTT CTGCCTCATG TGCAATAAG	3780

645

CTAATTCGTT GTGATGGTGT AAATATTCCA GAATTATTAG ATTCCATTCT GTATTTAATG	3840
CCATTAGATA GTTACGTCGA TAGTTCAATT CAGTTTATGA ACGTTGTTTC GGGTGATGAT	3900
ATTCCTAAGA TATGGGGTAC CTATAGACAG ATGATTGAAG GTCATGGTAC AGATCTTAAA	3960
ACGATTACTT ATCTTAGAAG AGAAGACTTT TATGAACGTA GTAAGAAAGC TTATGCTATT	4020
GTTGCTACAG GAGAAACTTC ACTTTATGCT AATATTATCC TTAAGAAAGG AGTAGTTGTT	4080
GAAAGAGAAA ATGTTCAATA GAGGAATTTT AGTTGCCAGT CATGGTAATT TTGCTAGCGG	4140
AGCTCTCATG ACCGCAGAAA TGTTTGTTGG TGAGACAACA AATGATAGAG TTAGGACATT	4200
AGGTTTGATG CCTGGAGACA ATATTGTAGA GTTTGAGCAT TATTTTAAAA ATCAAGTGG	4260
TGAAGTGTTA GACTCAAATC AAGAGGTTAT CGTTTTGACT GACTTGATTG GAGGAAGTCC	4320
TAATAATGTG GCTTTGTCAC GGTTTTTAAA TTTGGATTCA GTTGATATTG TAACAGGGTT	4380
TAATATCCCT CTCCTAGTGG AATTAATATC AAGTTATGAT TCAAAAATCA ATTTAGAAGA	4440
AATGTTCAC AATGCTCAAA ATAGTTTGTT TAATGTAAA CAACAACTTA ACGTAGAGGA	4500
GGAAGAAGAT TTATGTCTAT AGAGTTTGTT CGTATTGATG ACCGTCTGGT ACATGGTCAA	4560
GTTGTCAC TA CGTGGCTAAA AAAGTATGAT ATTGAGCAAG TTATCATTGT TAATGATCGC	4620
ATCTCAGAAG ATAAAACACG ACAATCTATT TTAAGATTT CTGCACCGGT AGGTTTAAAA	4680
ATTGTTTTCT TTAGTGTAAG ACGGTTTGTG GAAGTTTTAA ACTCTGTGCC AATAAAAAAG	4740
AGACAATGC TGATATATAC AAATCCAAA GATGTGTATG ATTCTATTGA AGGAAATTTA	4800
AAATGGAGT ACCTCAATGT AGGACAGATG ACTAAAACGG AGGAAATGA AAAGGTAACG	4860
GCAGGTGTAG CTCTAGGTGA AGAAGACAAA TATTATTTTA AGAAAATAGT TGATAAGGGA	4920
ACGAGAGTTG AAATTCAAAT GGTTCCTAAT GATAAAGTTA CAATGTTGGA AAAATTTTTA	4980
TAAAAATAAT TTAAGGAGGT ACAGTATATG CTATTCACAC AAGCATTACT GGTGACATTA	5040
GTTGGGATTA TTGCCACTAT TGACTATAAT GGACCGTTAT TTATGATTCA CCGTCCGTTA	5100
GTACAAGTG CAATGGTTGG CTTAGTATTA GGAGATTTCA CCCAAGCTGT TCTTATTGGT	5160
TCAGCTCTTG AATTAAGTTG GCTCGGTGTA ACAGGTATTG GAGGTTATAC TCCACCAGAT	5220
ACTATTTTCTG GTGCGATTAT TGGTACTGCA TTTGGTATTT TATCTGGTCA AGGAGAAACT	5280
GCTGGTATCG CTATAGCAGT TCCAATTGCA GTTGCTACCC AACAGTTGGA TGTCTTGCA	5340
AAAACTTTAG ATGTTTATTT TGTGAAAAA GCTGATAATG ATGCTAAAA CGGAGATTAT	5400
TCAAAGATCG GTTTTTATCA TTATTCAAGT TTGGTTTTAA TCACGTTATT TAAAAATTGA	5460
CCAAATTTCC TAGCTATTAT GCTTGGAGGG GAATATGTGG CAGACTTGTT TGCTAAGGTT	5520

646

CCACCAATCG TTATGCAGGG ACTTAACTCT GCAGGTGCTT TACTACCTTC AATTGGTTTT	5580
GGTATGCTTT TAAATATGAT GCTCAAGAAA AATATGTGGG TATTCTTGTT GATTGGATTC	5640
ATTTGTTCTG TGTATGGAGG AATGTCAACC ATTGGGATCT CACTAGTTGG TATTGCGGTA	5700
GCATACTTCT ACGATATGAT TGAAGCAAA CCACAAGAAA CAACTTCAAG TAGTGATGTT	5760
GAGGAGGATC TTGATCTATG ATGAATAATA AAGTAACTAA AGTTGAACTT AAAAAAGTTT	5820
TCAAACGAAG TTTTATGTAT GGTCTTCAT GGAAGTATGA GAGAATGCAG AACCTAGGTT	5880
TTCTATATAC AATCTTCCA GTATTGAAAA AACTATACCC AGACAAAGAT TCAGCTTCTC	5940
CTGCAATGAA ACGTCACCTT GAGTTTTCA ATACTCATCA AACAGCGGCA CCATTTATTC	6000
TTGGAGTTAC TTCCGCTATG GAAGAACAAG AAGGAAATGA AGGTGCAGCT TCAATTACTG	6060
GTATTAAAGT TGGCTTGATG GGGCCACTGG CTGGTCTAGG AGATAGTTTG TTCTGGCTGA	6120
CACTAGTTCC TATCTGTTTT AGTATTGGTG CGTCTTATTC TAAAGACGGC GGTGCTTTAG	6180
GTATCTTTAT CGCCTTAATA TTGTTAATA TTATTAATAT TCCTGTTAAA TATTTGCGTT	6240
TGAAATATGG GTATACTAAG GGTCTAGTC TTATCCAAGA AAATAATACA AAAGGAACAT	6300
TGAATCGCGT TACGAGTATG GCGACAGCAT TAGGGCTACT ACTAGTGGGT GGTTTGATTC	6360
CATCAATGGT TGGTATTAAT TTTGGATTAG AATTTAAGCA GGGGGAAGTT GTTATTTCTG	6420
TTCAAGAAAT GATTACAAAA TTAATCCAG GATTTATCCC TATGGCTTTG ACTTTATTAA	6480
TGTGTAAATT AATTAGAAAA GGAAAGAATC CGGTTGTACT AATCTTTAGT GTTATGGCTA	6540
TTGGAGTTAT TCTAGTTGTT TTAGGAATTT TGAAGTAGTA GAAAGTGTGG AGGTGGTATT	6600
TGGGATATCA CCTCCATTTT GGAAGAGAGG TAAAGAGTGA AATTATGGTA TAAGAAAGCT	6660
GGCGCAAATT GGAATGAAGC CTTGCCGATT GGGAACGGTC ATTTAGGTGG TATGATTTAT	6720
GGTTCAGCTA CAAAAGAATG TATTCAACTA AACGATGAGA CTATTTGGTA TAGAGGAAAG	6780
TCAGATAGAA ATAATCCAGA CTCACTATTG CATCTTAAAA AAATTCGGGA ATATCTTTTA	6840
GATGGAGAAA TTCAGAAAGC CGAAGAATTG ATAAAGTTAA CAGTGTGTC TACCCCAAGA	6900
GATCAAAGCC ACTATGAATT ACTTGGGGAA CTTTACATTG AGCATATAGA TATTCAGTCT	6960
TGTGCTCTTT CATTGTATGA AAGAGAGCTA GATTTAGATA CAGCTATTTT TAATGTTGTG	7020
TTTGAGCCTA ATAGTTGTAA TTTACAAATA AAAAGAGAAT ATTTTACGAG TTTTAATAAG	7080
AATATTTTAT GTTGCCGTAT AGTGTCATCA GTTCAAAACA CATTAAATTT AAACATTAAT	7140
TTGGGTAGAA ATAAACGGTT TAATGACGAA GTATCTAAAC TGGATTCAAG TACAATTTTA	7200
ATGTCGGCCT CTGCTGGAGG TAGAAAAGGT GTTCAGTTTA AAGTAGTATG TCATTCTAAG	7260
GTTACGGATG GTGAAGTAAG TGTATTGGGA GAGACAATAG TTATTCGGAA TGCTACAGAG	7320

647

GTATTTCTTT ATCTCAAATC AATGACGGAT TATTGGGGAA ATATAGATAT TTCTTCTCTT	7380
CAGGGAGAAT TTAGTAGTAT TGATTACTTT ACAGAAAAAG ATGAACATGT AAAAAAATAT	7440
CAGGAGCAAT TTAATAGAGT TGATTTTAAA CTAGACTATA GTAAAGGTTG TCTTAGCATT	7500
CCAACGAATC TACTTCTTGA AAACACTAAA AAGTATAGTA ACTACTTGAC TAAC TTGTTA	7560
TTTCATTATG GAAGATATCT GTTAATATCG TCTAGTCAAC CGAATGGTTT ACCTGCCAAT	7620
CTTCAAGGAA TATGGTGTGA TGAATTAAAT CCAATTGGG GTTCTAAATA TACGATTAAT	7680
ATTAATACTC AAATGAATTA TTGGATGGTA GGTCCATGTG ATTTACCAGA AGTAGAATAT	7740
CCATTATTG ATATGCTCGA AAGAATGAGA GAACCGGGAA GACTAACCGC TAAGAAAATG	7800
TATGGAGCTA GAGGTTTTAC AGCACATCAT AATACGGATG GTTTTGGCGA TACGGCTCCC	7860
CAATCTCATG CCATGGGGGC TGCAATTTGG GTATTAACTA TTCCATGGT ATGTACTCAT	7920
ATTTGGGAAC ACTATTTATA TTTCCAAGAT GAGCGTATTC TTACGGAACA TTTTGAAATG	7980
ATAAAGAAG CATTTCTTTT CTTTGAAGAT TATTTATTTG AGGTGGATGG CTAATTGATG	8040
ACAGGTCCAA GTGTCTCACC GGAAAATAAA TATCGCTTAA AAAATGGTAT TGAAGGAAAT	8100
GCTTGTCTAT CATCTACAAT TGATAATCAA ATTCTAAGAT ATTTTGTGA TTCATGCATT	8160
GGCATTGCAA AACAATTAGG AGACAATTCG GATTTTATTA GTCGTGTGAA GGAGTTAAAA	8220
AAGAACTAC CTAANAACAAA AATAGGTAGT AATGGGCAAA TCCAAGAATG GTTAGAAGAT	8280
TATGAAGAAG TAGAGCCTGG GCATAGACAC ATTTACCTC TATTTGGGCT TTATCCTTAT	8340
AATGAGATTG ATATTCATAA AACTCCGGAA TTAGCAGAAG CAGCTAAAAT CACTATCAAT	8400
AGGAGATTAT CAAACGCTAA TTTTATCT TCACAGGAGA GGGAGCAAGC GATTAATAAT	8460
TGGTTAGTAA GTGGTTTGCA TGCTAGTACA CAAACAGGTG GGAGTGCTGC ATGGCTGATT	8520
CATTTTTTTG CGAGACTATA TCAAGGTGAA CCTGCTTATA ACCAGATTAA TGGTTTGTTA	8580
AATAATGCGA CTCTTGCGAA TTTATTTCTT GACCATCCAC CATTTCAAAT TGATGGTAAT	8640
TTAGGTTTGG TGAGTGAAT TTGTGAATTA TTAGTACAGA GCCATCATAA TTGGTTATCA	8700
CTAATCCAG CTTTACCTTC TGCTTGGTCA GAAGGAGAAG TGAAAGGTTT CAGAGTAAGA	8760
GGAGGATATA AGGTATCGTT TGCTTGGAAA AATGGGGATA TAACATTCCT AAAATTGGAA	8820
GGAGGAAACA AAGATCAAAA AGTAAGAGTA AGAATATATG GCAAAAATAC TGATGTACAA	8880
AATATTGAAT TGGTATTTAA TTCAGAAAAA ATTATTGAGT TAAATTTTAA GGTATAAGTC	8940
ATGAATAAAG AAAAAATAAA AAGAAAATTA ATCACAATAT TGTTTGTATG TATTGGGATG	9000
TTATGTTTTG GATTGTTAGC AGGAGTTAAG GCTGATAATC GTGTTCAAAT GAGAACCACG	9060

648

ATTAATAATG AATCGCCATT GTTGCTTTCT CCGTTGTATG GCAATGATAA TGGTAACGGA	9120
TTATGGTGGG GGAACACATT GAAGGGAGCA TGGGAAGCTA TTCCTGAAGA TGTAAAGCCA	9180
TATGCAGCGA TTGAACCTCA TCCTGCAAAA GTCTGTAAAC CAACAAGTTG TATTCCACGA	9240
GATACGAAAG AATTGAGAGA ATGGTATGTC AAGATGTTGG AGGAAGCTCA AAGTCTAAAC	9300
ATTCCAGTTT TCTTGCTTAT TATGTGGGCT GGAGAGCGTA ATACAGTTCC TCCAGAGTGG	9360
TTAGATGAAC AATTCCAAAA GTATAGTGTG TTTAAAGGTG TTTTAAATAT TGAGAATTAT	9420
TGGATTTACA ATAACCAGTT AGCTCCGCAT AGTGCTAAAT ATTTGGAAGT TTGTGCCAAA	9480
TATGGAGCGC ATTTTATCTG GCATGATCAT GAAAAATGGT TCTGGGAAAC TATTATGAAT	9540
GATCCGACAT TCTTTGAAGC GAGTCAAAAA TATCATAAAA ATTTGGTGTT GGCAACTAAA	9600
AATACGCCAA TAAGAGATGA TCGGGGTACA GATTCTATCG TTAGTGGATT TTGGTTGAGT	9660
GGCTTATGTG ATAACGGGG CTCATCAACA GATACATGGA AATGGTGGGA AAAACATTAT	9720
ACAAACACAT TTGAAACTGG AAGAGCTAGG GATATGAGAT CCTATGCATC GGAACCAGAA	9780
TCAATGATTG CTATGGAAAT GATGAATGTA TATACTGGGG GAGGCACAGT TTATAATTTC	9840
GAATGTGCCC CGTATACATT TATGACAAAT GATGTACCAA CTCCAGCATT TACTAAAGGT	9900
ATTATTCCTT TCTTTAGACA TGCTATACAA AATCCAGCTC CAAGTAAGGA AGAAGTTGTA	9960
AATAGAACAA AAGCTGTATT TTGGAATGGA GAAGGTAGGA TTAGTTTCAAT AAACGGATTT	10020
TATCAAGGAC TTTATTGAA TGATGAAACA ATGCCTTTAT ATAATAATGG GAGATATCAT	10080
ATTCTTCCTG TAATACATGA GAAAATTGAT AAGGAAAAGA TTTCATCTAT ATTCCCTAAT	10140
GCAAAAATTT TGAATAAAAA TAGTGAGGAA TTGTCTAGTA AAGTCAACTA TTTAAACTCG	10200
CTTTATCCAA AACTTTATGA AGGAGATGGG TATGCTCAGC GTGTAGGTAA TTCCTGGTAT	10260
ATTTATAATA GTAATGCTAA TATCAATAAA AATCAGCAAG TAATGTTGCC TATGTATACT	10320
AATAATACAA AGTCGTTATC GTTAGATTG ACGCCACATA CTTACGCTGT TGTAAAGAA	10380
AATCCAAATA ATTTACATAT TTTATTGAAT AATTACAGGA CAGATAAGAC AGCTATGTGG	10440
GCATTATCAG GAAATTTTGA TGCATCAAAA AGTTGGAAGA AAGAAGAATT AGAGTTAGCG	10500
AACTGGATAA GCAAAAATTA TTCCATCAAT CCTGTAGATA ATGACTTTAG GACAACAACA	10560
CTTACATTAA AAGGGCATA TGGTCATAAA CCTCAGATAA ATATAAGTGG CGATAAAAAAT	10620
CATTATACTT ATACAGAAAA TTGGGATGAG AATACCCATG TTTATACCAT TACGGTTAAT	10680
CATAATGGAA TGGTAGAGAT GTCTATAAAT ACTGAGGGGA CAGGTCCAGT CTCTTTCCCA	10740
ACACCAGATA AATTTAATGA TGGTAATTTG AATATAGCAT ATGCAAAACC AACAACACAA	10800
ACTTCTGTAG ATTACAATGG AGACCCCTAAT AGAGCTGTGG ATGGTAACAG AAATGGTAAT	10860

649

TTTAACTCTG GTTCGGTAAC ACACACTAGG GCAGATAATC CCTCTTGGTG GGAAGTCGAT	10920
TTGAAAAAAA TGGATAAAGT TGGGCTTGGT AAAATTTATA ATCGCACAGA TGCTGAGACT	10980
CAACGTCTAT CTAATTTTGA TGTGATTCTA TATGACAATA ATAGAAACGA AGTTGCTAAG	11040
AAACATGTGA ATAATTTGTC GGGTGAATCT GTTAGTCTAG ATTTCAAAGA AAAAGGAGCA	11100
AGGTATATTA AAGTTAAATT ACTAACGAGT GGAGTGCCTT TGAGTTTAGC AGAAGTAGAG	11160
GTTCCTTAGAG AATCAGATGG TAAGCAATCT GAAGAGGATA TAGATAAAAT AACAGAAGAT	11220
AAAGTAGTCT CTACAAATAA GGTAGCTACT CAAAGTTCAA CCAATTATGA GGGTGTAGCT	11280
GCTTTAGCAG TTGATGGTAA TAAAGATGGA GATTACGGAC ATCATTCCGT GACTCATACT	11340
AAGGCAGATT CTAACGCTTG GTGGCAGGTC GATCTGGGAG AAGAGTTTAC GGTTCCTAAA	11400
GTTGATATTT ATAATAGAAC AGATGCCGAA CCTCAGCGTT TATCTAATTT TGATGTTATT	11460
TTTCTATCTT CATCAGGAGA AGAAGTTTTT AGAAGACATT TTGATAAAGT AGTTGATGGT	11520
TTGTTATCTT TAAAAGTACC TTCTGTAGGG GCTAAGCTAG TCAAAATAGA ATTAATCA	11580
GCAGCTATTC CGTTAAGTTT AGCGGAAGTT GAAGTCTATG GTTCAAAGAG AACTCCGAAG	11640
AAACTTTCTA ATATTGCATT AACAAAAGAA ACTCGACAGA GTTCAACGGA TTACAATGGT	11700
TTTTCTCGTC TAGCAGTTGA TGGAAATAAA AACGGAGATT ATGGTCATCA TTCAGTGACT	11760
CATACCAAAG AAGATTCTCC TTCATGGTGG GAGATAGATT TAGCACAAAC CGAAGAATTA	11820
GAAAAGTTAA TTATTTATAA TAGAACAGAT GCTGAAATTC AGAGATTATC AAATTTTGAT	11880
ATTATTATAT ATGATTCAAA TGATTATGAA GTTTTACAC AACATATTGA CAGTTTAGAA	11940
AGCAATAATC TATCCATAGA CTTAAAAGGA CTGAAGGGAA AAAAGGTTAG AATTTCTTTG	12000
AGAAGCGCAG GAATTCCTTT AAGTTTAGCA GAGGTAGAGG TTTATACTTA TAAGTAATTT	12060
TAAAAATTAT CACCCAGGCT ACCGTAAATA TAATGGAGAT GGTAGTATGA AAGAAACAGA	12120
AAAATAAGAG GAAAATAGTA TGATTCAACA TCCACGTATT GGGATTCGTC CGACTATTGA	12180
TGGTCGTCGT CAAGGTGTAC GCGAATCACT TGAAGTGCAA ACAATGAACA TGGCTAAAAG	12240
TGTGGCAGAT TTGATTTCAA GCACATTGAA ATATCCAGAT GGGGAACCTG TGGAAATGCGT	12300
GATTTCTCCA TCTACTATTG GCCGTGTACC AGAGGCTGCA GCTTCCCATG AGTTGTTTAA	12360
AAAATCAAAT GTTTGCGCAA CAATTACAGT TACACCATGC TGGTGTTATG GTAGTGAAAC	12420
TATGGATATG TCTCCAGATA TTCCTCATGC TATTTGGGGA TTTAATGGGA CAGAACCCCC	12480
AGGAGCTGTC TATCTTGACG CTGTACTAGC TTCACATGCT CAAAAAGGGA TTCCAGCCTT	12540
TGGGATTTAT GGAAGAGATG TTCAGGAAGC TAGTGACACA GATATTCCAG AAGATGTCAA	12600

650

AGAAAACTT TTACGCTATG CGCGTGCAGC TCTTGCAACT GGCTTGATGA GAGACACTGC	12660
TTACCTATCA ATGGGTAGTG TTTCGATGGG GATTGGTGGT TCTATTGTAA ATCCGGATT	12720
CTTCCAAGAA TACTTAGGAA TGCGAAATGA ATCGGTAGAT ATGACGGAGT TCACGCGCCG	12780
TATGGACCGT GGTATTACG ACCCTGAAGA GTTCGAACGT GCGCTCAAAT GGGTGAAAGA	12840
AAACGTAAAA GAAGGATTCG ACCATAACCG TGAAGACCTT GTTTTAAGCC GTGAAGAAAA	12900
AGATAGACAA TGGGAATTTG TTATTAAGAT GTTCATGATT GGACGTGACT TAATGGTTGG	12960
TAACCCAAGA CTTGCTGAAC TTGGTTTGA GGAAGAAGCG GTTGGTCACC ATGCTTTAGT	13020
AGCTGGTTTC CAAGGTCAAC GTCAGTGGAC AGACCATTTT CCAAATGGGG ACTTTATGGA	13080
AACTTTCCTC AATACTCAGT TTGACTGGAA TGGTATTCGA AAACCATTTG TATTTGCGAC	13140
AGAGAATGAT TCACTAAATG GTGTGTCTAT GCTCTTTAAT TATCTATTAA CAAATACTCC	13200
ACAAATCTTT GCTGATGTGC GTACTTATTG GAGCCCAGAG GCTGTTAAAC GTGTAACGGG	13260
ACATACTTTA GAGGTCGTG CTGCAGCTGG CTTCTTACAT CTAATCAACT CTGGTCTCTG	13320
TACATTGGAT GGTACAGGTC AAGCTACTCG AGATGGCAAA CCTATTATGA AACCATTCTG	13380
GGAGTTGGAA GAAAGTGAAG TGCAGGCTAT GCTTGAAAAT ACAGACTTCC CACCAGCAAA	13440
CCGCGAATAC TTCCGTGGAG GAGGATTCTC AACTCGTTTC TTGACGAAGG GGGATATGCC	13500
AGTAACAATG GTACGTCTCA ATCTTCTAAA AGGGGTTGGT CCAGTGCTAC AAATTGCAGA	13560
AGGTTACACA CTTGAACTTC CTGAAGATGT TCACCATACT TTAGATAATC GTACAGATCC	13620
AGGATGGCCA ACTACTTGGT TTGCTCCACG TTTGACAGGA AAAGGTGCTT TCAAGTCTGT	13680
CTATGACGTC ATGAATAATT GGGGAGCTAA TCACGGAGCC ATAACATATG GACACATTGG	13740
AGCAGACTTG ATTACCTTGG CTTCTATGTT GAGAATTCTT GTCAATATGC ATAATGTACC	13800
TGAGGAAGAT ATCTTTAGAC CTA AAAAATTG GTCCTTATTT GGAACAGAAG ATCTAGAATC	13860
AGCAGACTAT CGTGCATGTC AGTTGTTGGG GCCACTACAT AAATAAACT TGT TTATATA	13920
GGAGGTGAAC TTACGTCCCT CCTATCCTTT TAAAAAGATT TGTTAAACAA TTCACAAA1A	13980
ATTGAAAACG AATACAAAAA GTAATATAAT GATGTTAAAT AGATAGCGCG GAGGCGCAGG	14040
AGGAAAATTA TATGGCTATA TTTTATGTTT CGGCAGTCAA CCTTATTGGA AAAGGTGTTG	14100
TAAATGAAGT GGGTCCTTAT ATCAAGGAAC TTGGCTATAA AAAGGCACTT TTGGTGACAG	14160
ATAAGTACAT CGAAGGCAGT GATATTTTAC CTAAGACTTT AAAACCACTG GATACAGAAG	14220
GAATCGAATA T	14231

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:



651

(A) LENGTH: 16995 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGTTCTCTTA ACTTTTTTAG GATGGCATT CCGCTCTCA GGTACTCATT TTCTGCTgAA	60
GACGTTCTAA TTCTGTCCTC TCTTCAGGTC TCGTTTTTGG CTTACGTCCC ATTTTAGGTA	120
CTCTCCCTCT TGTTTTCTCA ACAATAGTAT ACCCGTTTTT CCTGTATTGT GCTAGCCAGT	180
TAAGAAGTAT CGTACGACTT GGGAGACCGT ATTCAAGAGA AACTCTATCT TTAGTCCAGC	240
CTTCATGTCA GACTTTATTA CTCATTTCTT GTTTTAAATC AGGAGAATAG TAACGATTTT	300
TTCTTTTTTT GACGAACTCT ATTCCGTAAC GATCAATCAA TTTAATCATG TACCTAATAT	360
TAGAATTGCT TATCCCAAAT TTATTGAAA GCTTCTCTAA GCTATATCCT TGTTTTCTAA	420
GTTCATAGAT CTGAACTTTA TCATCATAAG TTAGTTTCAT AATAAAAACA CCCCCAAAGT	480
TAGATTTTTT CTGTCTAACT TTTGGGGTGT AGTTCATGTA CACCTGATAT GATGCGTTTT	540
ATAATTTTTA AGCCTTTTTG CCCAGCCTCG TCAAAAGTAA TGTTTTGACA CAAAATCTGT	600
GACAAAACCT TAGTTTTAAA GGTTTTTAAC TTTGTATATA CTAGTTTTAA GAAAAGGAGG	660
ATGATCTAAT GGAAGAAAAA GTATCATTGA AAGTCAGGGT TCAAAAACCTA GGGACATCGC	720
TTTCAAATAT GGTATGCCCC AATATTGGAG CATTTATTGC TTGGGGAGTA TTGACTGCCC	780
TCTTTATCGC TGATGGCTAT CTGCCAAATG AACAGTTAGC TACTGTTGTT GGTCTATGT	840
TAACGTATTT ATTGCCAATC CTGATTGGTT ACACAGGTGG ATATATGATC CATGGCCAAC	900
GTGGTGCCGT TGTAGGAGCT ATTGCTACTG TTGGTGCAAT CACAGGTTCT AGTGTTCCTA	960
TGTTTATCGG AGCTATGGTA ATGGGCCCCAC TGGGAGGATG GACTATCAAG AAATTTGATG	1020
AGAACTTCCA GGAAAAAATT CGTCCCGGAT TTGAAATGTT AGTTAATAAC TTCTCAGCTG	1080
GTCTCGTTGG TTTTGCATTA TTGCTTTTGG CTTTCTACGC AATCGGTCCA GTCGTATCGA	1140
CTCTTACTGG AGCTGTTGGG AATGGTGTG AGGCTATTGT CAATGCTCGC CTCCTTCCTA	1200
TGGCTAATAT TATCATCGAA CCGGCTAAAG TCCTTTTCCT CAATAATGCC CTCAATCATG	1260
GCATTTTAC TCCTCTGGGA GTAGAACAGG TAGCTCAAGC TGGTAAGTCA ATTCTCTTCC	1320
TATTGGAAGC TAATCCTGGA CCAAGTCTGG GAATCTATT AGCTTATGCT GTATTGCGTA	1380
AAGGTTCTGC TAAATCTTCT TCTTGGGGGG CAATGGTTAT TCATTTCTTC GGAGGGATTC	1440
ATGAAATTTA CTTTCCTTAT GTTATGATGA AGCCTACTCT ATTTTATAGCT GCTATGGCAG	1500

652

GAGGTATCTC TGGAACCTTT ACTTTTCAAC TCTTAGACGC TGGTCTTAAA TCTCCAGCTT	1560
CACCAGGTTC TATTATTGCG ATTATAGCTA CGGCGCCAAA AGGTGTTTGG CCCCATCTAA	1620
ATGTTCTTTT AGGTGTTTTA GTGGCAGCAG TTGTTTCTTT CCTTGTAGCA GCCCTTATTC	1680
TTCATGCAGA CAAGTCAACT GAGGATTCGC TCGAAGCTGC TCAGGCGGCT ACCCAAGCAG	1740
CTAAGGCTCA GTCTAAAGGT CAGTTAGTAT CAACTTCTGT TGATGCAGTT GTTTCGACAG	1800
ACTCAGTGGA AAAAATCATT TTCGCCTGCG ATGCTGGTAT GGAAGCTCT GCTATGGGAG	1860
CTAGTATTCT TCGAGATAAG GTTAAAAAAG CAGGTCTAGA GATTCCAGTA TCTAATCAGG	1920
CAATCTCAA TTTGCTTGAT ACACCAAAAA CATTAATTGT TACTCAGGAA GAACTGACAC	1980
CAAGAGCTAA AGACAAGAGT CCAAGTGCTA TTCATGTTTC TGTGATAAT TTCTTAGCGT	2040
CCTCTCGTTA TGATGAAATT GTAGCTTCAT TAACAGGAGC TTCTCCAATA GCAGAAATTG	2100
AAGGAGATAT ACCAACTTCA GCACCAGTAG ATAGTCAGGA AAGTGACCTT AACCATATTG	2160
ATGCTGTAGT AGTTGCTTAT GGTAAAGCAC AGGGAAGTGC AACTATGGGC TGTGAAACGA	2220
TTCCGGCTAT TTTAGAAAC AAGAATATTC GTATTCCAGT TTCTACTGCC AAAATTTTCT	2280
AATTAGGTGA ATTTAATTCT AAAAACATAA TGATTGTAAC AACTATTTCT TTACAGGCAG	2340
AAGTGCAGCA AGCAGCACCG AATTCTCAAT TTCTTATTGT GGATAGTTTA GTAACAACAC	2400
CAGAATATGA CAAAATGGCT GCTAGAAATGT ACAAATAGAA CTAGAGGTTT CTAATTTACG	2460
AATGCTATTA ACCAAACGAG AAGAACAATT ATTGAAGGCT TTCCTACATG TAGGGAAGCT	2520
TTCAATGCAA GATATGACTG AAATCTTACA GGTTTCATCT AGAACAATTT ATCGAACTTT	2580
ATCAGATTG ACAGATAGCA TGGAGCAATA TGGAAATCGAA ATAACGAAGC ATGGGAAATA	2640
CTATATTTTG ACTGGAGAGT TGGATGATTT GCCGACAGAA CTTGAAGTGT TAGTTGAGTA	2700
TAGTCCCCAA GAAAGACAAG AGTTGATTAC CTATCGCCTT CTGACTGAGA GTGGTTTTGT	2760
CACCAATGAA GCATTGCAAG AGTGCACGAA AGTCAGTAAT GTAACATTA TTCAGGATAT	2820
TTCAGATATT GATAAGCGTC TTTTAGACTT TGATCTGAAA ATTGAACGAC AAAAAGGTTA	2880
TCGGATTCTT GGTGATTCAG TTGGTAAGAG AAGATTTTTG GCTATTTTAC TGACAAACTG	2940
TATCTCAGTA GCAGATTTTT CAACCGGTAA TTTTGGGAGC TTTGATATTT TAGAAGCAGA	3000
TAGAACTGGG CTGGCCAGTC AGATTGTAA TAAGCAACTG TCAGGTTTTT CAGATATGGA	3060
TGCTAGGATG AAGATGTTTT TTGCGATCTT GTTATCTCTT ATAGGTCAGG AGCAAAACAT	3120
TGAAAAATCA CCTAATACTA GTAAGCAGGC TTTGGAAATT TCTCAAAAAA TTTTTCAGC	3180
TTACTCTAAG CAGACTGCAC AATTTTATAG TATTTCAGGAA ATTATCTATT TTGCGAGCAT	3240
CTTGATGAA TTAATCATT AACGTCAGGA CAATCCGCTC TTTACGGAGA AATTGATGG	3300

653

TGAATTTTTC TACAATATTT CAAATCTGAT TGATACGGTT TCCATGTATA CCAAGATTGA	3360
CTTTTTTAAG GACAAGGTTT TATTCAATTT TCTTTTCCAT CATATTCCGC TCAGTTTAGG	3420
CGTCCCTATC CTTTTTCAGG GTGAAAATTT GCCAGAATCT ATCCAGATTT TAGTTGAAAG	3480
GAATAAATTT CTTTATACAG TCATCAGTCT TTTAGTGAAT GATATTTTTC CGAAATATCT	3540
TCATACAGAG TATGAGTATG GCATGATTGC CCTACATTTT ATCTCTAGCT TAGGCCGTAG	3600
TCCAGAGATT TATCCAGTCC GTGTTTGTCT TTTAACGGAT GAACGTCGGG TCACTAGAGA	3660
TTTATTAGTC AGTAAATTA AGAGTGTTC TCCTTTTGTG GAGTTGATAG ATATTCAATC	3720
TCTAGTAGAT TACCACAGTA TTGATCTCAG TCAGTATGAT TATATTTTAT CTACCAAGCC	3780
GCTGACTAAT CAGGAAATCG ATGTAATTTT TAGTTTTCCA ACCGTCAAAG AATTGCTTGA	3840
ATTACAGGAA CGACTTCAGT ATGTACAGGC ACATCGTACA ATTGTCGCGC GTGATGCTAT	3900
CGCTCCAGAG AAAAGTTATG ACTTGCAAGA TTATTTAATA TCTAGTAGTC AGCTTTTGAG	3960
TCAATTCGAG TTGGTTCAAT TGGAGAATAA TCAATCATTT GAGCACACGG TAGAACAAAT	4020
CATCCAATAT CAGAAGAATG TGAGTGACAG AGCTTACCTA ACAAGAAAAT TGTTATCTCA	4080
CTTCCAGAAT AGTCCTATGG CTATTCCTAA TACTGGTCTG GTGCTTTTAC ATAGTCAGTC	4140
TAGCAAAGTA ACAACAAATA GTTTTACTAT GTTTGAACTC AAACCTACCTA TCTCCGCATT	4200
GTCAATGAAA CGAGAGGAAG AAGAGGTCAA AAGGTGTCTG CTAATGCTAA TGTCTAAAGA	4260
AGCTAGCGAG GAAGCTAGAG ATTTAATGAC AGCTATTAGT CAGTCGATTA TTGAAAATCA	4320
TCTTTATACA GAGATTTACA AGACGGGAAA TCAATCCATT ATTTATCAGA TGCTAAATAC	4380
TATTTTTAAC GAAAAAATTA AGAAATTGGA GAACTAATAT GAACTTGAA AAACATTTGA	4440
TTAAGCTTAA TAAACAATTT TCTAACAAGG AGGAAGCTAT TTGTTATTGT GGGCAAGTTC	4500
TTTATGAGGG TGGATATGTT AATGAAGACT ATATTGAAGC CATGATTGAG CGAGATAAAG	4560
AGCTATCTGT TTACATGGGT AACTTTATCG CCATACCGCA TGGAACAGAT GCAGCAAAAA	4620
ATGATGTCCT CAAGTCTGGT ATTACAGTCG TTCAAGTCCC TAGAGGGGTT GATTTTGGGA	4680
ATGTATCTAA CCCTCAAGTG GCAACGGTTC TTTTGGTAT TGCTGGTATT GGTAATGAAC	4740
ACTTAGAAAT TATTCAGAAA ATTTCTATCT TCTGTGCAGA TGTAGATAAT GTTCTTAAAC	4800
TAGCAGATGC TCAGTCAAAA GAGGAAGTAT TGCGCTTATT TGATGCTGTT GAATAATTGA	4860
ATTTAGTCAT TTGTCATCTA GTATATATGT CCCTCAAATA GGAAAAGGAG AAATTGAATG	4920
AAACATTCTG TTCATTTTGG TGCCGGTAAT ATCGGTCGTG GTTTTATAGG TGAAATTCTA	4980
TTTAAAAATG GTTCCATAT TGATTTTGTG GATGTCAATA ATCAGATAAT TCATGCTCTG	5040

654

AATGAAAAGG GCAAGTATGA AATTGAAAT GCACAGAAAG GACAGTCTCG TATAGAAGTA	5100
ACTAATGTGG CTGGCATTAA TAGCAAAGAA CATCCTGAGC AAGTCATTGA AGCGATTCAA	5160
AAGACGGATA TTATTACTAC TGCAATCGGA CCTAATATAC TCCCTTTTAT CGCCGAACTT	5220
CTAGCCAAAG GAATCGAAGC TCGCCGAGTT GCAGGAAATA CACAGGCATT GGATGTTATG	5280
GCCTGTGAAA ATATGATTGG CGGGTCTCAA TTTCTTTATC AAGAAGTCAA GAAATATTTA	5340
AGTCCGGAAG GTTTGACATT TGCTGATAAC TACATAGGTT TTCCAAATGC TGCAGTAGAC	5400
AGGATTGTTT CAGCACAAAG TCACGAAGAT TCCCTTTTGT TTGTGGTCTG GCCCTTTAAT	5460
GAATGGGTCT TGGAAACCAA GCGTCTTAAA AATCCAGATT TACGTCTAAA AGATGTGCAT	5520
TATGAAGAAG ATTTAGAACC CTTTATTGAG CGAAAACTTT TTTCAGTCAA TTCTGGACAT	5580
GCAACTTCAG CTTACATTGG TGCGCATTAT GGTGCCAAGA CAATTTTGA AGCTCTTCAA	5640
AATCCTAATA TTAAATCTCG GATTGAATCT GTATTAGCTG AAATTCGGAG TCTCTTGATT	5700
GCCAAATGGA ACTTTGATAA AAAAGAATTG GAGAATTATC ACAAAGTCAT TATAGAACGA	5760
CTTGAAAACC CTTTCATAGT GGACGAGGTT AGTCGCGTAG CTCGTACTCC AATCCGAAAA	5820
TTAGGCTATA ATGAACGATT CATCCGCCG ATACGTGAAT TGAAAGAACT CAGTTTGTCA	5880
TATAAAAACC TACTTAAAC AGTTGGCTAT GTCTTTGACT ATCGCGATGT AAATGATGAA	5940
GAAAGTATTC GATTAGGTGA ATTGTTGGCT AAACAATCAG TCAAAGATGT TGTATACAA	6000
GTTACAGGTT TAGACGACCA AGAATTGATT GAGCAAATTG TAGAGTATAT TTAATCTTTT	6060
TCGAAAATCT CTTCAAATCA GGTTAGCATC GCTTTGTCTT AGGCATATGT TGTCTATCT	6120
ACAACCTCAA AGCAGTGCTT TGAGCTGACT CCGTCAGTCT TATCTGCAAT CTCAAAACAC	6180
TGTTTGAGTT ATCTGCGGTA ATCTTTCTAG CTTGTCTTTG ATTTTGTGTG TTATTTATAA	6240
GGTAAAAGAA GCTCGACAAA AAGTCTTCAA AATCGGGAAA AGGCAGCCTA TCGGGTGTTC	6300
AAAAATCTTG ATAGGATGTC CTTTATTATG GAAAGCCTTA TTGGATTTTC TCCTCAGATT	6360
GAGTTTTTGA TCAGCTTTAT GAGATAGGTC TTGCTAGAGA TGTAGCCCAT CATGTTATT	6420
TTATGGACAG TGGGAAAATT GTTGAAAAAA ATAATGCCCA TCAATTCTTT AGTCGTCCAA	6480
GAGAAGAACG AACCAAGCAA TTTTGAACG AATTCTTTCT AATGCGATCT ATATAGTAAA	6540
ATGAAACAAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTCTTAA AAATGTTTTA	6600
GAAGTAGAGG TGTACTATTC TAGTTTCAAT CTAATATATA ACTGAAAAAT TAGATAAATT	6660
AGTTTTGGAA AATGACTAAC CAAAAGATAT CCAAAGTAGT CTAAAATTGT CTATACTTTA	6720
TGAGTGTTTT AGTTAGGAAA AAGGCTTGTT GTCTATAATT GTCTGCATTA GTCTAGATTT	6780
TATTTATAGA AAATGTTATA ATAGACTGTA TTTAAAAAAT TTTAAGGAGA AATGACAGAA	6840

655

TGTCTGTATC	ATTTGAAAAC	AAAGAAACAA	ACCGTGGTGT	CTTGACTTTC	ACTATCTCTC	6900
AAGACCAAA	CAAACCAGAA	TTGGACCGTG	TCTTCAAGTC	AGTGAAGAAA	TCTCTTAATG	6960
TTCCAGGTTT	CCGTAAAGGT	CACCTTCCAC	GCCCTATCTT	CGACCAAAAA	TTTGGTGAAG	7020
AAGCTCTTTA	TCAAGATGCA	ATGAACGCAC	TTTTGCCAAA	CGCTTATGAA	GCAGCTGTAA	7080
AAGAAGCTGG	TCTTGAAGTG	GTTGCCCAAC	CAAAAATTGA	CGTAACTTCA	ATGGAAAAAG	7140
GTCAAGACTG	GGTTATCACT	GCTGAAGTCG	TTACAAAACC	TGAAGTAAAA	TTGGGTGACT	7200
ACAAAAACCT	TGAAGTATCA	GTTGATGTAG	AAAAAGAAGT	AACGTACGCT	GATGTGGAAG	7260
AGCGTATCGA	ACGCGAACGC	AACAACCTGG	CTGAATTGGT	TATCAAGGAA	GCTGCTGCTG	7320
AAAACGGCGA	CACTGTTGTG	ATCGACTTCG	TTGGTTCTAT	CGACGGTGTT	GAATTTGACG	7380
GTGGAAAAGG	TGAAAACCTC	TCACCTGGAC	TTGGTTCAGG	TCAATTCATC	CCTGGTTTCG	7440
AAGACCAATT	GGTAGGTCAC	TCAGCTGGCG	AAACCGTTGA	TGTTATCGTA	ACATTCCCAG	7500
AAGACTACCA	AGCAGAAGAC	CTTGCAGGTA	AAGAAGCTAA	ATTCGTGACA	ACTATCCACG	7560
AAGTAAAAGC	TAAAGAAGTT	CCGGCTCTTG	ACGATGAACT	TGCAAAAGAC	ATTGATGAAG	7620
AAGTTGAAAC	ACTTGCTGAC	TTGAAAGAAA	AATACAGCAA	AGAATTGGCT	GCTGCTAAAG	7680
AAGAAGCTTA	CAAAGATGCA	GTTGAAGGTG	CAGCAATTGA	TACAGCTGTA	GAAAATGCTG	7740
AAATCGTAGA	ACTTCCAGAA	GAAATGATCC	ATGAAGAAGT	TCACCGTTCA	GTAAATGAAT	7800
TCCTTGGGAA	TTTGCAACGT	CAAGGGATCA	ACCCTGACAT	GTACTTCCAA	ATCACTGGAA	7860
CTACTCAAGA	AGACCTTCAC	AACCAATACC	AAGCAGAAGC	TGAGTCACGT	ACTAAGACTA	7920
ACCTTGTTAT	CGAAGCAGTT	GCCAAAGCTG	AAGGATTTGA	TGCTTCAGAA	GAAGAAATCC	7980
AAAAAGAAGT	TGAGCAATTG	GCAGCAGACT	ACAACATGGA	AGTTGCACAA	GTTCAAAACT	8040
TGCTTTCAGC	TGACATGTTG	AAACATGATA	TCACTATCAA	AAAAGCTGTT	GAATTGATCA	8100
CAAGCACAGC	AACAGTAAAA	TAATCTTAAT	AAACAGAAAA	CCCACCTGAA	TTGGTGGGTT	8160
TTCTGATGCA	CTATTTTCCA	AAAATCTCTT	TGAGGTCTGT	GTCTGTAATC	CCAATCATGG	8220
CTGGGATGCG	GTCCCAGTTT	TCTTCGGTTA	GGATGTAGGA	TTGTTTCAGAG	GCACTTGATG	8280
TGACTGTTTC	AGAGACAGCT	TGTTGCTTTT	CTTCAACATT	CTCCAGTAGA	TCACTGAAGC	8340
GTTCAATCAG	ATAGGTTTTT	CGGGCAGTTC	CGATGTGTTG	GGTAGCATAG	TCGAAGGCTT	8400
GTAATTCGCC	TAGTAAGATG	AGTTTGCTTT	TGGCACGTGT	AATGGCTGTG	TAGATGAGAT	8460
TTCCGCTCCAG	CATACGTCGG	CTAGCACTAG	TAATCGGTAG	GATGACAACT	GGGAACCTCAC	8520
TTCCCTGAGA	CTTATGAATA	CTCATGGCAT	AGGCCAAGCG	AATCTTGTA	CATTCTGTTAC	8580

656

GGGGTAAGA GACTTCATTA CCATCAAAAT CAATGACAAT CTCGTCTTGT TTCGATTCGG	8640
TGTATTTACC AGGAATCAGG TCTGTGATAG CTCCTAAATC CCCATTAAAG ACATTGATTT	8700
CAGCATCGTT AACCAAATGA ATGACCCTGT CTCTCTTACG ATAGTGACAC TGAGGAGCTT	8760
CAAAACTGAG TTGATCTTTT TGTGGGGGAT TGAGCAGGTC TTGCATGAGC TGATTGATAG	8820
CATCAATCCC TGCCGTCCCT CGGTACATAG GAGCCAGAAC TTGGATATCA CGGGCGGGAA	8880
TACCATTTCT GAGGGCGGCA CCTAAGATTT TTTCAATCGT GGCAGGAATA TGGCCACTAG	8940
CAATTTCAAA GTAGGAACGG TCAGCTTTTT TTTGGGTGAA ATCAGCTGGC AAGATGCCCT	9000
GTCGAATCTG ACTAGCTAGG GTGACGATGG TTGATTCTTT GCTTTGTGCA TAAATTTTTT	9060
CCAAGCGAGT CTGAGGAATC AAAGGAATAT GAAGTAGATC CGCTAGAACC TGTCCAGGAC	9120
TGACAGAAGG TAGCTGATCA CTGTACCTA CGATGAGGAT CTTACTGTTA GAAGAGATAT	9180
TGGAGAAGAG TTGATTGGCC AGCCAAGTAT CTACCATAGA GAATTCATCC ACGATGATAA	9240
AGTCAGCATC TAGGTAATCT TCCAGATGAC TGGTATCATC GTCACCTGTC ATTCCCAAGT	9300
GGCGATGTAT GGTGCGGCTA GGCAAACCTG TCAATTCATT CATGCGACGA GCAGCTCGAC	9360
CAGTTGGAGC AGCAAGAAGA ATGGGCAGAT TGCTTTTCTT CCTGAAGTCA AGTCCTTCTA	9420
AAAGGGCATA AACAGCAATG ATTCCATTGA TAACAGTTGT CTTACCAGTA CCAGGCCCCAC	9480
CTGTGAGGAT AAAGACCTTA TTCTGGATAG CATCACAGAT AGCCTGTTTT TGAATGTTAT	9540
CATACTCAAT TCCCAGTTCT TGCTCGACAG TAGTGATATG TTTTGAATG GTTTCTAAAT	9600
CATGACTCTT CTGTTTTCTT TTTTCAAGGA TACGAACCAA GTGACTGCGG ATGCCCTTCT	9660
CAGCGAAAAA GAGGCTGTTG TCAAAGATCT TGGTATCAAT CTGCTGAACC TTGTCTTCTT	9720
CGATCAGGTA GGAGAGCTCT TGGGCAACTT GGCTGGGGTC TAGTCCACG GGACGGGAAG	9780
ACTCAAGGAG AGTAAGGGTT TGTTCAGCA AATCCCGTGC TTCAACATAG GTGTCCCCTG	9840
TTTCCATACA GGCTGAAAA AGACTGTGAA CTAGACCGGC GCGGAAGCGT TCAGGAGCCT	9900
GACTTTCGAT GCCTAGTTCC TCAGCTAGTT GGTGAGCAAT GGTAAAGCCC AAACCCTTGA	9960
TATCCTCAAC CAGCTGGTAG GGATAATTTT CAACCACATC AAGGGTTTCT TCCTTGTAAT	10020
AGTCTTGAAT CTGAAAGGCT AGTTTGTGG GAATGCCGTA GTTGGCTAGT TTGGCCAAAA	10080
TCATCTCCGT TCCGTAGTTG AGACGGAGAG TGGAGACGAA AGCCTCGCGA TTTTGGCAG	10140
AGAGTCCCTGC GATGCCTTCT AACTTTTCTG GGTGTTGCAA AATTTCTGTA ATGGTATTTT	10200
CGCCATAGGT ATCCACGATT TTCTGAGCTG TCTTGAGACC AATCCCTTG AAATGGCTAC	10260
TTGAAAAGTA CTTGACCAAG CCCTTACTAG TTGGTTTTGC GCGATCATAA CACTGATTT	10320
GCAGTTGTTT TCCATACTTG GAGTGCTGGA CAATTTGCCC CAAAAAGTA TAGTCTTCGC	10380

657

CCTCAATTAC ATCAGCCATG GTTCCTGTGA CAATGATTTC AAAATCATCA AAATCCTCTG	10440
CGTCCGTATC GTCGATTTCCT AGGAGGAGGA TGCATAAAA ATTGCTGGGA TTTTCAAAAA	10500
TAATCCGTTC AATAGTTCCT GAAAAATAAA CTTCCATAAA ATTCCTTTGC ATGAATAGGT	10560
GAGAGTTGGG ATTGTTTTTA TTTTATACTC TTCGAAAATA TCTTCAAACC ACGTCAGCTT	10620
CCATCTGCAA CCTCAAAACA GTATTTTGAG CTGACTTCGT CAGTTCTATC CACAACCTCA	10680
AAACACTGTT TTAAGCAGCC TACGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTCATTGAG	10740
TATTTGTAAG TAAACAATCA CTTCTCAGCA TAGAAGAAGA GGCTGAGATT GGTGATTCTC	10800
TGCCTCTTAG GTTCTTAAAT ATGTTCCGAT ACGGGTGATT GGCCATAAGC GGAATTAGC	10860
TTCCCTGTG ATATCTTTTG CTTTGAAGGT ACCTACGTGG CGGCTGTGCG TCGAAACCAA	10920
GCGGTCATCT CCGAGGAGAA GGTATTCTCC TTCTGGAACA GTAAAGCTAA AGTTGGTGTT	10980
GTAGTTGACA TCAACTGTGA AGGCTTGAGC TTTTGTAGCG ATACTTCTAA AGAAAGTTCC	11040
TTTATTTCTT TCAAAGCCCT TGCCTGAGTA AGTGCTTTGG AGTTTGTCTT CTTTGAAGCG	11100
TTTGATATAG TCTGCTAGAT AAGGCTCGTC CGTTTCTTTG TCATTGATGT AGAGTTTATC	11160
ATTTTCGTAA CGAATGGTGT CGCCAGGCAT TCCAATCAGC CGCTTGACGA TGTCTTATT	11220
GCCATCTTCC TCATGGGCCA CCACGATATC AAAACGGTCA ATAGGAAGGT GTTTTACAAC	11280
GAAGAGAATT TCGCCATCCG CTAGGGTCGG ATCCATGGAA TGTCTTCTA CGCGAACATT	11340
GCTCCAAAAA AAGATACGAC TTAAAGCTAG TAATGACAGA ATTAGGAGGA ACAATCCCCA	11400
CTCTTTTAAG AAATTTTAA ATGAATTCAT AACTTACCTT TCTAAGCGTT TTTTCGCTTT	11460
TTAGTGTTT TTAAAGTGCA ATTTGGCGCA GAAGCTGACT CCCTGCATAC CATAGGCTTG	11520
CAAAATCTGG CTAGCCACCT TGTGAGAAGC CGTTCCAGCT CCACTTGGGA GCTGATAACC	11580
CAGTTCTCGT CCCAAATTTT CAAGATTTTC CAGAAAGAGA TCACGCGCAA TGACAGAAGA	11640
AACTGCGACA GACAAGTATT TGCCCTCAGC CTTTCTTCTT AAGCTGATAG GATTGCTGAA	11700
ACGATTGGCC TCTTGTGCCA AGTACTTGTC ATAATTTTPTA GCACTGGTAA AGGCATCAAT	11760
CACAATTTTC TCAGGCTGAA CACCTTTTGT AAGGAGGAGA TAGATAGCCT GATTATGGAG	11820
GGCAACCTTA ACCGAAACAG CGTTGTAGCG GTCTCCGATG ACCTCGTTGT ACTTGCTGGG	11880
TGAGAGAAGG AGTGCTGCTT GCTGAATTTT TTCCTTGAGA ATAGGAGTAA TCTGACGGAT	11940
CTTTTGGTCG GTCAGAGTCT TAGAATCCCC CACACCGAGT TTTCTGAAAA AGTCGTGCTG	12000
GTCAGGTGTG ACAAAGGCAG CCACAACGTC AAGCCCACCA AAGTAGGAAC CATTTCCAC	12060
CTCATCTGTC CCAATTAAAG GAAGATTTTG TCCGCTGGTT TGCTCTACAG CTTGATAGCC	12120

658

AAAGAACTG GCGTATTTT CAGCCCCTC ACCCTGAAGC AAGATTTTC CAGAAGTATA	12180
GATAGAAACC GTTGCTTGAG GTAGTTTCAA AAAGTAGCGG ATATAGGGAT TCTTGCTAGG	12240
AGCCAGACTG GTTTGATAGT GTTCAAGAAA AGCCTGAATA TCCTTTTCGC TTGGTGTGAG	12300
TGTGATACTT GCCATAGTTT CTATTGTACC ACAAAGCAG TAAAATTTGT AAAAAGTAC	12360
AAAATTAGCG AATTTTGCTA TAATATCGTG AGGTGAATTT TATGGCAAAT CTAAATCGAT	12420
TCAAATTTAC ATTCGGGAAA AAATCGTTAA CCTTGACAAG CGAACATGAC AACCTTTTTA	12480
TGGAGGAAAT CGCTAAGGTT GCGACAGAAA AATACCAAGC AATTAAAGAA CAAATGCCTA	12540
GCGCAGATGA TGAAACAATC GCTCTTTTGT TGGCAGTCAA CTGTTTATCA ACTCAGCTCA	12600
GCCGTGAGAT TGAATTTGAC GATAAGGAGC AAGAGCTAGA AGAACTCCGT CACAAGCTTG	12660
TGACTTGTA GCAAGAACAG AGCAAGATTG AGGATTCTTT ATGATTTCAT TCCTTCTTCT	12720
ATTGGTCTTG GTTTGGGGAT TTTATATCGG CTATCGGAGA GGCCTGCTCT TACAGGTTTA	12780
TTACCTGATT TCAGCCATGG CATCGGCTTT TATGGCTGCC CAGTTTTATA AGGGGCTTGG	12840
AGAGCAATTC CATTTATTGC TCCCTTATGC AAATTCGCAG GAAGGTCAGG GGACTTTCTT	12900
TTTCCCATCG GATCAACTCT TTCAGCTGGA TAAGGTCTTT TATGCAGGTA TCGGCTACTT	12960
GCTTGATTTT GGGATTGTCT ATAGCATTTG TCGTTTACTT GGTCTTCTCT TACACTTGAT	13020
TCCTAGCAAA AACTGGGTG GTAAGTTGTT CCAAGTTTCA GCAGGTATCT TGTCCATGTT	13080
GGTGACCTTA TTTGTCTTGC AAATGGCCTT GACAATCTTG GCGACCATCC CCATGGCAGT	13140
TATACAAAAT CCTCTTGAAA AGAGTATCGT CGCAAAACAC ATCATCCAGA GCATACCGGT	13200
AACAACCACT TGGCTCAAAC AAATCTGGGT GACAAATTTA ATCGGATAAA AAGGGCAGGA	13260
GTTTTCTAG CCCTTTGTTT ACAGATTTGA CTCGAATCTA TCAGAATGTA AAAAGTACC	13320
ACACCTAGAC ATTCAAAGAC AAGGAAATAA AGATGAATAA GAAAATATTA GAAACATTAG	13380
AGTTCGATAA GGTCAAGGCC TTGTTTGAGC CTCATTTGTT GACCGAGCAG GGCTTGAGC	13440
AATTGAGACA ACTGGCTCCG ACTGCCAAAG CAGATAAAAT CAAACAGGCT TTTGCTGAGA	13500
TGAAGGAAAT GCAGGCTCTT TTCGTCGAGC AACCGCATTT TACTATTCTC TCAACTAAGG	13560
AAATTGCAGG AGTCTGCAAG AGGTTGGAGA TGGGAGCGGA TCTCAATATC GAGGAGTTCC	13620
TACTCTTGAA ACGCGTGCTT CTTGCCAGCC GAGAACTTCA AAATTTTAC ACCAATCTGG	13680
AAAATGTCAG CTTGGAAGAA TTAGCCCTTT GGTGTGAGAA ATTACATGAT TTTCCGCAAT	13740
TACAAGGAAA TCTTCAGGCC TTTAATGATG CGGGTTTCAT TGAAAATTTT GCCAGTGAAG	13800
AATTGGCGCG AATCCGTCGA AAAATACATG ATAGCGAGAG TCAGGTACGC GATGTTTAC	13860
AAGACTTGCT CAAGCAAAAA GCGCAGCTGT TGACGGAAGG AATTGTTGCT AGCAGAAATG	13920



559

CCCGTCAGGT TTTACCAGTC AAAAACACCT ACCGCAATAA GATTGCAGGT GTCGTTTCATG	13980
ATATTTCTGC TAGTGGAAC ACCGTCTATA TCGAACCCCG TGAGGTAGTC AAAGTGAAGC	14040
AAGAAATTGC TAGTCTGCGA GCAGATGAGC GCTATGAAAT GCTTCGCATT CTCCAAGAAA	14100
TTTCTGAGCG TGTCCGCCCT CATGCGGCTG AGATTGCTAA TGACGCTTGG ATTATCGGTC	14160
ATCTGGACTT GATTCTGCC AAGGTTGAT TTATCCAAGA AAGACAAGCA GTCGTGCCTC	14220
AGCTGTCAGA AAATCAAGAG ATTCAACTGC TCCATGTCTG CCATCCTTTG GTCAAAAATG	14280
CCGTCGCAA TGATGTCTAT TTTGGTCAAG ATTAAACAGC TATTGTCATT ACAGGTCCCA	14340
ATACAGGTGG GAAGACCATC ATGCTCAAAA CTCTGGGCTT GACACAGGTC ATGGCCCAGT	14400
CAGGATTGCC GATTTTAGCA GACAAGGGAA GTCGTGTTGG TATTTTTGAA GAAATCTTTG	14460
CTGATATTGG AGATGAGCAG TCTATTGAGC AGAGCTTGTC TACCTTCTCT AGTCATATGA	14520
CCAAATATCGT GGATATTCTT GGCAAGGTCA ACCAACATTC ACTCTTACTT TTGGATGAGT	14580
TGGGGGCTGG TACTGATCCC CAAGAGGGAG CAGCCCTTGC CATGGCTATT CTGGAGGACC	14640
TTGCGCTGCG TCAAAATCAAG ACCATGGCGA CGACCCACTA TCCAGAACTC AAGGCCTACG	14700
GTATTGAGAC AGCCTTTGTG CAAAATGCCA GTATGGAGTT TGATACTGCA ACTCTTCGCC	14760
CGACCTATCG CTTTATGCAG GGTGTTCTG GCCGAAGTAA TGCCTTTGAA ATTGCCAAAC	14820
GTCTAGGCCT ATCTGAAGTT ATCGTAGGAG ATGCCAGTCA GCAGATCGAT CAGGACAATG	14880
ACGTCAATCG TATCATTGAG CAATTAGAAG AGCAGACGCT GGAAAGCCGC AAACGTTTGG	14940
ACAATATCCG TGAGGTGGAG CAAGAAAATC TCAAGATGAA CCGTGCCTA AAAAACTCT	15000
ACAACGAGCT TAATCGTGAA AAGGAAACCG AGCTTAACAA GGCGCGTGAA CAGGCTGCTG	15060
AGATTGTGGA TATGGCCCTA AGTGAAAGTG ACCAGATTCT CAAAAATCTC CACAGTAAAT	15120
CCCAACTCAA GCCCCACGAA ATCATTGAAG CCAAGGCCAA GTTGAAAAAA TTGGCTCCTG	15180
AAAAAGTGA CTTGTCTAAA AATAAGGTCC TTCAAAAGGC CAAGAAAAAA CGAGCTCCAA	15240
AGGTGGGAGA TGATATCGTG GTTCTCAGTT ATGGTCAGCG TGGTACCTTG ACCAGTCAAC	15300
TCAAGGACGG TCGCTGGGAA GCCCAAGTTG GCTTGATTAA GATGACCTTG GAAGAGAAAG	15360
AGTTTGATCT TGTTCAAGCC CAGCAAGAAA AACCAGTCAA GAAGAAACAG GTCAATGTTG	15420
TGAAACGAAC TTCTGGGCGA GGACCTCAAG CTAGACTGGA TCTTCGAGGC AAGCGCTATG	15480
AAGAAGCCAT GAATGAGCTA GATACCTTCA TCGACCAAGC CTGCTTAAAC AATATGGCTC	15540
AAGTTGATAT CATCCATGGT ATCGGAACAG GAGTCATCCG TGAAGGAGTT ACCAAATACT	15600
TGCAAGAAA CAAACATGTC AAGAGTTTCG GCTATGCCCC ACAAATGCT GGAGGCAGTG	15660

660

GTGCGACTAT TGTCACTTTT AAAGGATAGC AGTATTCTGG ACTTTATAAA GTAAAAACTG	15720
TTGAACTAAT TTTTACTAAT AAACACATTG ACAAAGCCA ACATTTTTTG TAAAATTAGA	15780
ATCAATTAAA TACCAACACC GAATGAAGTT TAATAGAAGT GGGGAATCGT TTGATTTTCC	15840
ATGACTGTAA ATGGACGGAA CTCTGGAGAG ACCGTAAAGG CACCGAAGGG CAAGGCAGGC	15900
AACTGCTCAA ACTCTCAGGT AAAAGGACAG AGCTAGGATA GATCGCTTTT TAGCATTTAT	15960
CTAAGCATTC CAGAGTACAT GTATCTTGCA TGTGCTCTTT CTTTTGGGGT TGAAACGATA	16020
GGAGAAGGAA ATGTTAGAAT TGCTTAAATC AATCGATGCT TTTGCTTGGG GACCGCCCCT	16080
CTTGATTTTA TTGGTCGGAA CAGGGATTTA CCTAACTATT CGCTAGGAC TCTTGCAGGT	16140
TTTGCGTCTA CCCAAGGCCT TTCAGCTTAT TTTTATCCAG GATAAGGGAC ATGGTGATGT	16200
ATCCAGTTTT GCAGCTCTGT GTACAGCCTT GGCATCAACT GTTGGACAG GAAATATCAT	16260
AGGAGTTGCG ACGGCTATCA AGGTTGGTGG ACCAGGAGCT CTATTTTGA TGTGGATGGC	16320
GGCTTTCTTT GGAATGGCTA CCAAGTATGC GGAAGGACTC TTGGCCATCA AATACCGCAC	16380
CAAGGACGAC CATGGTGCAG TAGCGGGAGG TCCCATGCAT TATATCCTTC TAGGGATGGG	16440
AGAAAAGTGG CGACCACTTG CTGTTTTGTT TGCAGTAGCA GGAGTATGG TTGCTCTCTT	16500
GGGAATCGGA ACCTTCACCC AAGTCAACTC GATTGCAGAA TCTATCCAAA ATACAACGAC	16560
GATTTTCGCCA GCCATCACAG CTCTCGTCTT GTCTGTCTTT GTAGCGATTG CAGTCTTTGG	16620
TGGACTCAAG TCTATTTCTA AGGTTTCAAC TACTGTTGTT CCTTTTATGG CCATCATTTA	16680
TATCTTAGGA ACTCTTACAG TTATTTTCTT TAATATCGGA AAAATCCCTG GCACAATCGC	16740
TTTAGTCTTT ACCTCAGCTT TTAGTCCCCT TGCTGCGGTA GGTGGATTG CTGGTGCTAG	16800
CGTTCCGATG GCTATTCAAA ATGGTGTGGC GCGTGGTGTG TTCTCAAACG AATCTGGTCT	16860
GGGTTCGCT CCTATTGCAG CTGCAGCTGC CAAGACAAAT GAACCAAGTAG AGCAAGGTTT	16920
GATTTCCATG ACAGGAACCT TTATTGATAC CCTCATCATT TGTACTCTAA CTGGTTTGAC	16980
CATCTTGGTA ACTGG	16995

## (2) INFORMATION FOR SEQ ID NO: 83:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28473 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCGGGGCTTT TGTAGTATAA TAGAGATACG TTTTGAAAGT AGGAGGTATC TATGGACTTA 60

661

ACTAAGCGCT TTAATAAACA GTTAGATAAA ATTCAAGTTT CGTTGATTGG TCAGTTTGAC	120
CAGGCTATTT CGGAGATTCC TGGGGTCTTG CGTTTGACCT TGGGGGAACC TGATTTTACA	180
ACGCCAGACC ATGTCAAGGA GCGGGCAAG CGAGCGATTG ATCAGAACCA ATCCTACTAT	240
ACAGGGATGA GTGGTCTGCT GACTCTACGT CAGGCAGCCA GTGACTTTGT TAAGGAAAAG	300
TACCAACTGG ACTATGCTCC TGAAAATGAA ATCTTGTTA CAATTGGGGC GACAGAGGCT	360
TTATCTGCGA CTTTGACGGC TATTTTGAA GAGGGAGACA AGGTACTTTT GCCAGCTCCT	420
GCTTATCCAG GCTATGAACC GATTGTTAAC TTAGTTGGGG CAGAAATGT TGAGATTGAT	480
ACGACTGAAA ATGGTTTTGT CTTGACTCCT GAGATGTTGG AGAAGGCCAT TTTGGAGCAG	540
GGTGATAAGC TCAAGGCGGT TATTCTCAAC TATCCAGCCA ATCCGACAGG AATTACCTAC	600
AGTCGAGAGC AGTTAGAGGC CTTGGCAGCT GTTTTACGCA AGTACGAAAT TTTTGTGTGTC	660
TGTGATGAGG TTTACTCAGA ATTGACCTAC ACAGGCGAAG CCATGTGTCT CTAGGAACGA	720
TGTTGAGAGA CCAGGCTATT ATTATCAATG GTTTGTCTAA ATCGCATGCC ATGACAGGTT	780
GGCGTTTGGG GCTGATTTTC GCTCCTGCGA CCTTCACAGC CCAGTTAATC AAGAGTCACC	840
AGTACTTGGT CACTGCCGCA AATACCATGG CGCAACATGC TCGGGTAGAA GCCTTGACGG	900
CTGGTAAAAA CGATGCCGAC CCATGAAGAA GGAATATATC CAACGTCGGG ACTATATCAT	960
CGAAAAATG ACTGCTCTTG GTTTTGAGAT TATCAAACCA GACCGTGCCT TCTATATTTT	1020
TGCTAAAAAT CCAGCGGGCT ACAATCAAGA CTCCTTTGCT TTTCTGAAGG ATTTTGCTCA	1080
GAAGAAGGCC GTTGCCCTTA TCCCTGGTGC AGCCTTTGGA CGTTACGGGG AAGGCTACCT	1140
CCGCTATCT TATGCAGCCA GCATGGAGAC TATCAAAGAA GCCATGAAAC GACTTGAGGA	1200
GTACATGAGA GAAGCATGAT TCAGTCTATC ACGAGTCAAG GCTTGGTGCT TTACAATCGC	1260
AATTTTCGTG AGGATGACAA GCTCGTCAAA ATTTTACAG AGCAGGTTGG CAAACGCATG	1320
TTTTTTGTCA AACACGCTGG TCAGTCTAAG CTGGCGCCTG TTATTGAGCC CTTGGTGCTG	1380
GCACGATTTT TCTTGCGAAT CAATGATGAC GGACTCAGTT ACATCGAAGA CTATCATGAG	1440
GTCATGACTT TTCCCAAGAT TAATAGTGAC CTCTTTGTCA TGGCCTATGC GACCTATGTG	1500
GCAGCTCTTG CAGATGCTAG TTTGCAGGAC AATCAGCAGG ATGCTCCCTT GTTTGCTTTT	1560
TTGCAAAAGA CTTTGAGATT GATGGAAGCA GGCTTGGATT ATCAGGTTTT GACCAATATT	1620
TTTGAAATTC AAATTTTGAC TCGATTTGGA ATCAGCCTCA ATTTTAATGA GTGTGTCTTC	1680
TGCCATCGGG TTGGTCAGGC TTTGACTTT TCTTTCAAAT ATGGAGCCTG CCTCTGTCCA	1740
GAGCATTATC ATGAGGATAA GAGACGTTGT CATCTCAATC CCAATATCCC CTATCTGCTC	1800

662

AATCAATTTC	AAGCTATTGA	TTTTGAGACT	TTGGAGACCA	TTTCGCTCAA	GCCTGGAATC	1860
AAGCAAGAGC	TACGCCAATT	TATGGATCAA	TTATATGAAG	AGTACGTTGG	GATTCACCTA	1920
AAATCAAAGA	AATTTATTGA	TTCCCTAGCA	GACTGGGGAC	AATTACTAAA	AGAGGAAAAG	1980
AAATGAAAAA	AATCGCAGTA	GATGCCATGG	GGGGCGATTA	CGCACCTCAG	GCCATTGTTG	2040
AGGGTGTCAA	TCAAGCCCTA	TCTGACTTTT	CAGATATCGA	GGTTCAACTT	TACGGAGATG	2100
AAGCTAAAAT	CAAGCAATAT	CTGACAGCGA	CAGAGCGCGT	CAGCATTATC	CATACGGATG	2160
AGAAGATTGA	TTCGGATGAT	GAACCTACGA	GAGCTATTCG	GAATAAGAAA	AATGCCAGTA	2220
TGGTATTGGC	AGCCAAGGCT	GTCAAAGATG	GTGAAGCAGA	CGCTGTCCTT	TCGGCTGGGA	2280
ATACAGGTGC	CTTGTTGGCA	GCAGGATTCT	TCATCGTGGG	TCGTATCAAG	AATATCGACC	2340
GTCCTGGACT	CATGCTTACC	TTGCCTACCG	TTGATGGAAA	AGGTTTTGAC	ATGCTAGACC	2400
TTGGTGCCAA	TGCAGAAAAT	ACAGCCCAGC	ACCTCCATCA	ATATGCGGTT	CTAGGTTCTT	2460
TCTATGCTAA	AAATGTCCGT	GGCATTGCGC	AACCACGCGT	TGGTTTGCTC	AACAACGGAA	2520
CAGAGAGTAG	CAAGGGCGAC	CCGCTTCGTA	AGGAAACTTA	TGAATTACTG	GCGGCTGATG	2580
AAAGTTTGAA	CTTTATCGGA	AACGTGGAAG	CGCGTGATTT	GATGAATGGC	GTTGCAGATG	2640
TTGTTGTGGC	AGATGGTTTC	ACGGGAAACG	CTGTGCTCAA	ATCCATCGAA	GGGACAGCTA	2700
TGGGAATCAT	GGGCTTGCTC	AAGACAGCTA	TTACAGGTGG	TGGTCTTCGA	GCGAAACTAG	2760
GTGCCCTCCT	TCTCAAGGAC	AGCCTCAGTG	GTTTGAAAAA	ACAGCTCAAT	TATTCAGATG	2820
TTGGTGAGC	GGTCTTGTTT	GGTGTTAAGG	CACCTGTTGT	CAAGACTCAT	GGCTCAAGCG	2880
ATGCCAAGGC	TGTTTATAGT	ACGATTCTGC	AGATCCGTAC	CATGCTAGAA	ACAGACGTGG	2940
TTGCCCAGAC	TGCCCGTGAA	TTTTCAGGAG	AATAAAGAG	ATGACAGAAA	AAGAAATTTT	3000
TGACCGTATT	GTGACCATTA	TCCAAGAGCG	ACAGGGAGAG	GACTTTGTCT	TGACAGAATC	3060
CTTGAGTCTG	AAAGACGATT	TGGATGCGGA	TTCTGTTGAC	TTGATGGAGT	TTATCTTGAC	3120
TCTGGAAGAT	GAATTTAGTA	TCGAAATCAG	CGATGAAGAA	ATTGACCAAC	TCCAAAACG.	3180
AGGAGATGTG	GTTAAAATCA	TTCAAGGAAA	ATAGCAATCG	GAGTTCCAAG	TCAACGGAAG	3240
TAGATGGTTT	TTAGAAATGA	GAAATATCGG	ACAAGCTGGT	AAAATCTTGG	CTGACAGTGG	3300
TTATCAAGGG	CTCATGAAGA	TATATCCTCA	AGCACAAACT	CCACGTAAAT	CCAGCAAAC	3360
CAAGCCGCTA	ACAGTTGAAG	ATAAAGCCTG	TAATCATGCG	CTATCTAAGG	AGATAAGCAA	3420
GGTTGAGAAT	ATCTTTGCCA	AAGTAAAAAC	GTTTAAAATG	TTTTCAACAA	CCTATCGAAA	3480
TCATCGTAAA	CGCTTCGGAT	TACGAATGAA	TTTGATTGCT	GGTATTATCA	ATCATGAACT	3540
AGGATTCTAG	TTTTGCAGGA	AGTCTAATAG	TAAAAAAGTG	ATTAGAAAAC	ATCTTTTTTA	3600

663

AAAATAGAGA TGATTTTGAA ACAAAAAAGC TAATTCAAGA CGTTTCGATG CCAATTCAAG	3660
ATTTGGATGA AAAAAATTAA TAGATACTGT TATACTAAAC TTGTCAAGTT TGTAACAAGA	3720
CAAATATTAA AAATAAAAAA GAGGTATTCG TTATGAATAC AAAAACGATG TCACAATTTG	3780
AAATTATGGA TACTGAGATG CTGCTTGCG TTGAAGGTGG CGGATGCAAT TGGGAGATT	3840
TTGCCAAAGC AGGTGTTGGA GGAGGAGCAG CACGAGGTCT TCAGCTAGGA ATTAAAACAA	3900
GAACATGGCA AGGTGCAGCA ACTGGTGCTG TGGGAGGAGC TATACTTGA GGTGTGGCCT	3960
ATGCAGCGAC ATGTTGGTGG TAATTATGGA TTTTAAAGT TTTATTATTG GTTTAGTAGT	4020
TGGTATATTT GGTCCCTTATA TGGATGATTT AATTAGAAAA AAATTTTAA AGTCTTCGGA	4080
GAAGAAAACA GAAAAATCTG TTAATAATA ATCAAACTA TAAATGATGA ATCTGAATCA	4140
AAATTATTTT GCGCATGTAA AGAGGAGTCT TATAGTAACG AGTCAAAAAA GGAGTAACTA	4200
TGAATCGTAA TTTAGAACGG TGTATCTAT TCTGACTAGG AATAGATCAT ACCAGAGGTA	4260
GCTTAGAAAT AGCAGAGACA TTAGAAATTG AAGTAATAAA TAGGATGTCG TAAGTGTTAC	4320
TATCAATGAT TTATTTGTTT CAAGCTTGCC TAGGGTGACA GTAAAAAATC AATTCCTTT	4380
CAATAGCATA TTTTLAGTGG GCAGGACTCT TGTCTGCCT ATTTTATTAT CCAAAAAGTG	4440
CAGTTGGGAG GGAGATAGGC TCATTGGGA AGGAAGTCCA GTTTTGTGT AGTGATTGGG	4500
GTAAGATAGT TGTATCAGA TGAGTTAATA CTCTCGAAA ATCAAATTCA AACCACGTCA	4560
ACGTCGCCCT GCCGTATATA TGTGACTGAC TTCGTCAGTC CTATCTACAA CCTCAAAACA	4620
GTGTTTTGAG CAGCCTACGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA	4680
GGGAAAAGGA GATGAATATG AAATTGGGA AACGTCATTA TCGTCCGCAG GTGGATCAGA	4740
TGGACTGCGG TGTAGCTTCA TTAGCCATGG TTTTGGCTA CTATGGTAGT TATTATTTT	4800
TGGCTCACTT GCGAGAATTG GCTAAGACGA CCATGGATGG GACGACGGCT TTGGGCTTGG	4860
TCAAGGTGGC AGAGGAGATT GGTTTTGAGA CGCGAGCCAT TAAGGCAGAT ATGACGCTTT	4920
TTGACTTGCC GGATTTAACT TTTCCTTTTG TTGCCCATGT GCTTAAGGAA GGGAAATTGC	4980
TCCACTACTA TGTGGTGACT GGGCAGGATA AGGATAGCAT TCATATTGCC GATCCAGATC	5040
CCGGGGTGAA GTTGACTAAA CTGCCACGTG AGCGTTTTGA GGAAGAATGG ACAGGAGTGA	5100
CTCTTTTAT GGCACCTAGT CCAGACTATA AGCCTCATAA GGAACAAAAA AATGGTCTGC	5160
TCTCTTTAT CCCTATATTA GTGAAGCAGC GTGGCTTGAT TGCCAATATC GTTTTGGCAA	5220
CACTCTTGGT AACCGTGATT AACATTGTGG GTTCTTATTA TCTGCAGTCT ATCATTGATA	5280
CCTATGTGCC AGATCAGATG CGTTCGACAC TAGGGATTAT TTCTATTGGG CTAGTCATCG	5340

664

TCTACATCTT CCAGCAAATC TTGTCTTACG CTCAGGAGTA TCTCTTGCTT GTTTGGGGC	5400
AACGCTTGTC GATTGACGTG ATTTTGTCTT ATATCAAGCA TGTTTTTCAC CTCCCTATGT	5460
CCTTCTTTGC GACACGCAGG ACAGGGGAGA TCGTGTCTCG TTTTACAGAT GCTAACAGTA	5520
TCATCGATGC GCTGGCTTCG ACCATCCTTT CGATTTTCCT AGATGTGTCA ACGGTTGTCA	5580
TTATTTCCCT TGTTCATTTT TCACAAAATA CCAATCTCTT TTTTCATGACT TTATTGGCGC	5640
TTCTTATCTA CACAGTGATT ATCTTTGCCT TTATGAAGCC GTTTGAAAAG ATGAATCGGG	5700
ATACCATGGA AGCCAATGCG GTTCTGTCTT CTTCTATCAT TGAGGACATC AACGGTATTG	5760
AGACTATCAA GTCCTTGACC AGTGAAAGTC AGCGTTACCA AAAAATTGAC AAGGAATTTG	5820
TGGATTATCT GAAGAAATCC TTTACCTATA GTCGAGCAGA GAGTCAGCAA AAGGCTCTGA	5880
AAAAGGTTGC CCATCTCTTG CTTAATGTCG GCATTTCTCTG GATGGGGGCT GTTCTGGTCA	5940
TGGATGGCAA GATGAGTTTG GGGCAGTTGA TTACCTATAA TACCTTGCTG GTTTACTTTA	6000
CTAATCCTTT GGAAAATATC ATCAATCTGC AAACCAAGCT TCAGACAGCG CAGGTTGCCA	6060
ATAACCGTCT AAATGAAGTG TATCTAGTAG CTTCTGAGTT TGAGGAGAAG AAAACAGTTG	6120
AGGATTTGAG CTTGATGAAG GGAGATATGA CCTTCAAGCA GGTTCAATTAC AAGTATGGCT	6180
ATGGTCGAGA TGTCTTATCG GATATCAATT TAACCGTTCC CCAAGGGTCT AAGGTGGCTT	6240
TTGTGGGGAT TTCAGGGTCA GGTAAGACGA CTTTGGCCAA GATGATGGTT AATTTTACG	6300
ACCCAAGTCA AGGGGAGATT AGTCTGGGTA GTGTCAATCT CAATCAGATT GATAAAAAAG	6360
CCCTGCGCCA GTACATCAAC TATCTGTCTC AACAGCCCTA TGTCTTTAAC GGAACGATTT	6420
TGGAGAATCT TCTTTTGGGA GCCAAGGAGG GGACGACACA GGAAGATATC TTACGGCGCG	6480
TCGAATTGGC AGAGATTGCA GAGGATATCG AGCGCATGCC ACTGAATTAC CAGACAGAAT	6540
TGACTTCGGA TGGGGCAGGG ATTTTCAGGTG GTCAACGTCA GAGAATCGCT TTGGCGCGTG	6600
CTCTCTTGAC AGATGCGCCG GTCTTGATTT TGGATGAGGC GACTAGCAGT TTGGATATTT	6660
TGACAGAGAA GCGGATTGTC GATAATCTCA TTGCTTTGGA CAAGACCTTG ATTTTCATTG	6720
CTCACCCTT GACTATTGCT GAGCGGACAG AGAAGGTAGT TGTCTTGAT CAGGGCAAGA	6780
TTGTCGAAGA AGGAAAGCAT GCTGATTTGC TTGCACAGGG TGGCTTTTAC GCCCATTGG	6840
TCAATAGCTA GAAAGAGGAG AGGATGAAAC CAGAAATTTT AGAAAGTGCG GAGTTTATA	6900
ATCGTCGTTA CCATAATTTT TCCAGTAGTG TGATTGTACC CATGGCCCTT CTGCTTGTGT	6960
TTTTACTTGG CTTTGCAACT GTTGCAGAGA AGGAGATGAG TTTGTCCACT AGAGCTACTG	7020
TCGAACCTAG TCGTATCCTT GCAAATATCC AGTCAACTAG CAACAATCGT ATTCTTGTC	7080
ATCATTTGGA AGAAAATAAG CTGGTTAAGA AGGGGATCT TTTGGTTCAA TACCAAGAAG	7140

665

GGGCAGAGGG TGTCCAAGCG GAGTCCTATG CCAGTCAGTT GGACATGCTA AAGGATCAAA	7200
AAAAGCAATT GGAGTATCTG CAAAAGAGCC TGCAAGAAGG GGAGAACCAC TTTCCAGAGG	7260
AGGATAAGTT TGGCTACCAA GCCACCTTTC GCGACTACAT CAGTCAAGCA GGCAGTCTTA	7320
GGGCTAGTAC ATCGCAACAA AATGAGACCA TCGCGTCCCA GAATGCAGCA GCTAGCCAAA	7380
CCCAAGCCGA AATCGGCAAC CTCATCAGTC AAACAGAGGC TAAAATTCGC GATTACCAGA	7440
CAGCTAAGTC AGCTATTGAA ACAGGTGCTT CCTTGGCCCG TCAGAATCTA GCCTACTCTC	7500
TTTACCAGTC CTACAAGTCT CAGGGCGAGG AAAATCCCCA AACTAAGGTT CAGGCAGTTG	7560
CACAGGTTGA AGCACAGATT TCTCAGTTAG AATCTAGTCT TGCTACTTAC CGTGTCCAGT	7620
ATGCAGGTTT AGGTACCCAG CAAGCCTATG CGTCAGGGTT AAGCAGTCAA TTGGAATCCC	7680
TTAAATCCCA ACACTTGGCA AAGGTGGTC AGGAATTGAC CCTTCTAGCC CAGAAAAATT	7740
TGGAGGCAGA GTCAGGTAAG AAGGTACAGG GAAATCTTTT AGACAAGGGG AAAGTTACGG	7800
CGAGTGAGGA TGGGGTGCTT CATCTTAATC CTGAGACCAG TGATTCTAGC ATGGTTGCAG	7860
AAGGTGCCCT ACTAGCCCAA CTTTATCCAT CTTTGGAAG AGAAGGGAAA GCCAAACTCA	7920
CAGCTTATCT AAGTTCAAAA TATGTAGCAA GAATCAAGGT CGGTGATTCT GTTCGCTATA	7980
CTACGACTCA TGATGCCGGG AATCAACTTT TCCTAGATTC TACTATTACA AGTATTGATG	8040
CGACAGCTAC TAAGACTGAG AAAGGGAATT TCTTTAAAT CGAGGCGGAG ACTAATCTAA	8100
CTTCGGAGCA GGCTGAAAAA CTTAGGTACG GGGTGAAGG CCGCTTGCAG ATGATTACGG	8160
GCAAGAAAAG TTACCTACGT TATTATTTGG ATCAATTTTT GAACAAAGAG TAATGTTCCG	8220
GTMTTLAGAG TTAAATAATT TTAAACTGT GAGAAAGATT CTTCTTGCAG TTTTTCTTT	8280
ACAAATTTTG AAAACATCT ACTATTTATT CGGTAAAT CTTGTGTTTT TTGGTTTTTT	8340
GTGGTAAAT GTGCTCAAGT AATACGAAAG GCGAACTTTA AAATGTCAA ACAATTGATC	8400
TATTCGGGAA AAGCTAAAGA TATCTATACA ACTGAGGATG AAAATCTTAT TATTTCAACT	8460
TACAAGGACC AGGCGACTGC TTTCAACGGT GTCAAGAAG AGCAGATTGC AGGTAAGGGA	8520
GTCTGAATA ATCAGATCTC ATCTTTTATT TTTGAGAAAT TAAATGTGGC TGGTGTGGCG	8580
ACTCACTTTG TGGAGAACT TTCAGACACG GAACAACTCA ATAAAAAGGT TAAGATTATT	8640
CCTTGGAAG TCGTGTCCG CAACTATACT GCTGGTTCCT TTTCAAAACG TTTTGGTGTG	8700
GATCAGGGAA TCGCCTTGA GACTCCGATT GTCGAATTTT ACTACAAAAA TGATGATTTG	8760
GATGATCCAT TTATCAATGA TGAGCATGTG AAATTCCTAC AGATTGCGGG TGACCAGCAG	8820
ATTGCCTACT TGAAGGAAGA AACGCGTCGT ATCAATGAAC TATTGAAAGT CTGGTTTGCT	8880

666

GAGATTGGGC TTAAATTGAT TGACTTTAAG CTAGAGTTTCG GTTTTGACAA GGATGGCAAG	8940
ATTATCTTGG CAGACGAATT TTCACCAGAT AACTGCCGCT TGTGGGACGC TGATGGCAAC	9000
CACATGGATA AGGATGTTTT CCGTAGAGGA TTGGGAGAAC TAACCGACGT TTATGAGATT	9060
GTTTGGGAAA AGTTGCAGGA ATTGAAATAA TCTGTTTGCA ACGGAAAACC TTCGTCTCTC	9120
AACTAAAAGG ACTCAGGCTG AAAAGGTCCC CCAGACCTTT TCACTCTGTA GAGAACTAGG	9180
TGAACTAACA GATGTTTACG AAATTGTCTG GGAAAAGTTG CAGGGTTTAA AATAACAACC	9240
TCAAGGCTGT TTGGGAATAT TGCAAGAGCT GAAATAAAGG AATAAGAATT GATGGATAAA	9300
CGTATTTTTG TTGAAAAAAA GGCTGATTTT CAGGTCAAGT CAGAGAGTTT GGTTAGAGAG	9360
CTCCAGCACA ACTTGGGACT GTCAAGCTTG AAAAGTATTC GTATTGTGCA AGTATATGAT	9420
GTATTTGACT TGGCTGAGGA CTTGTTTGCA CCTGCAGAGA AGCACATTTT CTCTGAGCAG	9480
GTAACCGACC ATGTTTTAGA TGAAGTATCT GTGCAGGCGG ATCTTGCTAA CTATGCTTTC	9540
TTTGCCATTG AAAGTCTGCC AGGGCAGTTT GACCAGCGTG CAGCTTCGTC ACAGGAAGCC	9600
TTGCTTTTGT TGGGAAGTTC GAGTGACGTG ACAGTCAACA CAGCCCAACT TTAAGTGGTG	9660
AATAAGATA TTGATGCGAC TGAGTTGGAA GCTGTCAAAA ACTACCTGCT CAATCCAGTT	9720
GATTCTCGTT TCAAGGATAT CACGACAGGG ATTGCCAAGC AGGAGTTTTT AGAGTCAGAC	9780
AAGACCATTG CCAAATTGAC TTTCTTTGAA AGCTATGCAG CAGAAGACTT TGCTCGCTAC	9840
AAGGCCGAAC AAGGGATGGC CATGGAAGTG GATGATTTGC TCTTTATCCA AGACTACTTT	9900
AAGTCAATCG GGCCTGTCCT AACTGAGACT GAACTCAAGG TTTTGGACAC TTAAGTGTCT	9960
GACCACTGCC GTCATACGAC TTTTGAGACA GAGTTGAAAC ACATCGACTT TTCAGCTTCT	10020
AAATTTCAAA AGCAATTGCA GTCAACCTAT GACAAGTATA TTGCCATGCG CGAGGAATTA	10080
GGTCGGTCTG AAAAACCACA AACCTTGATG GATATGGCGA CTATTTTCGG TCGTTATGAG	10140
CGTGCTAATG GACGATTGGA TGATATGGAA GTCTCTGACG AAATCAATGC CTGCTCAGTT	10200
GAAATTGAAG TGGACGTTGA TGGTGTCAAG GAACCTTGGC TCCTCATGTT TAAAAACGAA	10260
ACCCACAACC ATCCAACAGA AATTGAGCCA TTTGGTGGAG CGGCTACCTG TATTGGTGGG	10320
GCTATTCGTG ATCCGTTGTC AGGCCGTTCC TATGTTTACC AAGCCATGCG TATTTAGGT	10380
GCTGGTGATA TTACAGCACC GATTTTCGGAA ACTCGCGCTG GGAAATTGCC ACAACAAGTC	10440
ATTTCTAAAA CAGCAGCTCA TGGTTATTCT TCATATGGTA ACCAGATTGG GCTTGCAACA	10500
ACCTACGTTT GTGAATACTT CCACCCAGGC TTTGTAGCTA AACGTATGGA ACTTGGTGCC	10560
GTTGTTGGTG CGACTCCCAA GGGCAATGTT GTCCGTGAAA AACCTGAAGC AGGTGATGTG	10620
ATCATCCTTC TCGGAGGCAA AACAGGTCGT GATGGTGTG GTGGTGGCAG GGGCTCTTCT	10680



667

AAGGTTCAAA CAGTTGAGTC TGTAGAGACT GCTGGTGGTG AGGTTCAAAA AGGAAATGCC	10740
ATCGAAGAAC GCAAGATTCA GCGCCTCTTC CGTAATGGCA ATGTCACCTCG TCTGATCAAG	10800
AAGTCCAATG ACTTTGGGGC AGGCGGTGTC TGTGTGGCTA TCGGTGAATT GGCAGACGGT	10860
CTTGAAATCG ACCTCAACAA GGTGCCTCTT AAATACCAGG GCTTGAATGG TACAGAAATT	10920
GCCATCTCTG AATCACAAGA ACGGATGGCG GTCGTGGTTC GTCCTGAAGA TGTGGATGCC	10980
TTCGTTGCCG AATGTAACAA AGAAAATATT GATGCTGTTG TGGTGGCGAC AGTAACTGAA	11040
AAACCAAATC TTGTCATGCA CTGGAATGGT GAGACAATCG TTGACTTGGA GCGTCGTTTC	11100
CTTGACACCA ATGGTGTGCG CGTGGTTGTC GATGCCAAAG TTGTGGACAA GGATGTCAAA	11160
CTCCAGAAG AGCGTCAAAC ATCTGCTGAA ACACTGGAAT CAGATACCCT TACGGTTCTA	11220
TCTGACCTCA ACCATGCAAG TCAAAAAGGA TTACAGACTA TCTTTGACTG CTCTGTTGGA	11280
CGCTCAACGG TTAATCACCC ACTTGGTGGT CGTTACCAAC TCACACCAAC TGAGGCATCT	11340
GTGCAGAAAT TGCCAGTTCA ACACGGTGTG ACTCATACTG CGTCGGTCAT TGCTCAAGGT	11400
TTCAACCCAT ATGTAGCTGA ATGGTCTCCA TACCACGGTG CTGCTTATGC GGTATCGAA	11460
GCAACTGCTC GTTTGGTGGC TGCTGGTGCC AACTGGTTCA AGGCTCGTTT CTCTTACCAA	11520
GAGTATTTTCG AGCGTATGGA TAAACAAGCA GAGCGTTTCG GTCAGCCAGT AGCTGCTCTT	11580
CTAGGTTCTA TTGAAGCACA AATTGAGCTT GGCTTGCCAT CTATCGGTGG TAAGGACTCC	11640
ATGTCTGGTA CCTTTGAAGA ATTGACCGTT CCGCCAACCT TGGCTGCCTT TGGGGTGACG	11700
ACGGCAGATA GCCGTAAGGT GCTCTCTCCA GAATTTAAAG CTGTTGGGGA AAATATCTAC	11760
TACATCCCAG GTCAAGCCCT CTCTGCAGAG ATTGATTTTG ACTTGATTAA GAAAAATTTT	11820
GCTCAGTTTG AAGCCATCCA AGCTGACCAT AAAGTGACAT CTGCATCAGC TGTCAAATAC	11880
GGTGGTGTAG TTGAAAGTTT GGCTCTTGCT ACCTTTGGAA ACTATATTGG TGCAGAGGTG	11940
ACCTTGCCCTG AACTTGAAAC AGCTTTGACA GCTCAATTAG GCGGCTTTCT CTTCACATCT	12000
CCTGAAGAAA TTGCTGGAGT AGAGAAGGTT GGACAAACGA AAGCAGACTT TACACTGACT	12060
GTCAACGGTG TGAAGCTAGA TGGACACAAG CTTGACAGTG CATTTCAAGG GACATTGGAA	12120
GAAGTTTACC CAACAGAATT TACCCAAGCG AAAGAACTAG AAGAAGTACC AGCTGTGGCA	12180
TCAGATGTTG TGATTAAAGC CAAAGAAAAG GTTGAAAAAC CTGTGGTTTA CATCCCAGTC	12240
TTTCCAGGAA CCAACTCAGA ATATGATTCA GCTAAGGCCT TCGAAAAAGA AGGTGCAGAG	12300
GTCAATTGCG TGCCATTGCT GACCTTGAAT GAAGAAGCTA TTGTCAAGTC AGTTGAAACT	12360
ATGGTTGACA ATATCGACAA GACTAATATT CTCTTCTTTG CTGGTGGATT CTCGGCTGCG	12420

668

GATGAACCAG ATGGTTCAGC TAAGTTTATC GTCAATATCC TGCTTAATGA AAAAGTGCCT	12480
GTGGCTATTG ATAGCTTTAT CGCCCGTGGT GGTTCGATTA TCGGTATTG TAATGGATTG	12540
CAAGCCTTAG TCAAATCGGG TCTCCTACCC TACGGAACT TTGAAGCTGC TAACAGTACT	12600
AGCCCAACCC TCTTCTACAA TGATGCCAAC CAACACGTGG CCAAGATGGT GGAAACTCGC	12660
ATTGCCAATA CCAACTCACC ATGGTTGGTT GGTGTGCAAG TGGGCGATAT CCACGCTATT	12720
CCTGTTTCGC ACGGTGAAGG GAAGTTTGTC GTGACGGCTG AGGAATTTGC AGAGCTCCGT	12780
GACAAATGGAC AAATTTTCAG CCAATACGTT GACTTTAACG GTAAACCAAG TATGGATTCT	12840
AAGTACAATC CGAATGGTTC TGTCCATGCC ATCGAAGGAA TTACCAGCAA GAATGGTCAA	12900
ATCATCGGTA AGATGGGCCA CTCAGAACGT TATGAGGATG GTCTTTTCCA AAATATCCCA	12960
GGCAATAAAG ACCAACACCT GTTCGCATCA GCGGTAAAC ATTTCACTGG AAAATAAGAC	13020
TTACAGATTT TCTAATAGAT AGTATCAGTA ATGTAAAAGT CATGTAAATC TAGCTCTTGA	13080
TGATTACAAA TGAAAATTAG GTATAAAAAA TGACATACGA AGTAAATCT CTTAATGAAG	13140
AATGTGGTGT TTTCGGTATT TGGGGACATC CAGATGCTGC TAAGTTGACC TATTTTGGAC	13200
TCCACAGTCT TCAACACCGT GGTGAGGAGG GGGCAGGAAT CCTCTCCAAT GATCAAGGAC	13260
AACTGAAGCG CCATCGTGAC ATGGGGCTTT TATCAGAAGT TTTCAGAAAT CCAGCTAATT	13320
TGGATAAATT GACAGGAGCT GGTGCGATTG GGCATGTGCG TTATGCGACT GCTGGCGAAG	13380
CTTCTGTAGA TAACATCCAG CCCTTCCTCT TCCGTTTTCA CGATATGCAG TTTGGTTTGG	13440
CTCATAATGG AAATCTGACC AATGCAGCCT CTCTCAAGAA AGAACTGGAA CAAAGAGGAG	13500
CAATTTTCAG CGCGACTTCG GACTCTGAAA TCTTGGCTCA CCTCATTCGT CGCAGTCATA	13560
ATCCTAGCCT GATGGGCAAA ATCAAGGAAG CGCTCAGCCT TGTCAAAGGT GGTTTTGCCT	13620
ATATCTTGCT GTTTGAGGAC AAGTTGATTG CGGCTCTTGA CCCAAATGGA TTCCGACCGC	13680
TTTCGATTGG TAAAATGGCT AATGGAGCAG TTGTTGTATC TTCTGAAACC TGTGCTTTTG	13740
AGGTCAATTG TGCCGAGTGG ATTCGTGATT TGAAGCCAGG TGAGATTGTG ATCATTGATG	13800
ACGAGGGCAT TCAGTATGAC AGCTATACAG ATGATACCCA GTTGGCGGTT TGTCTATGG	13860
AGTATATCTA CTTTGCTCGC CCTGATTCTA ATATCCACGG TGTCAATGTC CATACGGCAC	13920
GTAAGAGAAT GGGAGCGCAA TTGGCGCGAG AATTTAAGCA TGAGGCAGAT ATTGTAGTTG	13980
GTGTGCCCAA TTCTTCCTA AGCGCGGCTA TGGGATTGTC GGAAGAATCA GGCTTACCAA	14040
ATGAAATGGG TCTGATCAAA AACCAATACA CCCAGCGAAC TTTTATCCAA CCGACTCAAG	14100
AATTGCGGGA GCAAGGAGTG CGGATGAAAC TGTCTGCTGT TTCGGGTGTT GTCAAAGGCA	14160
AACGTGTGGT CATGGTGGAT GATTCCATTG TACGTGGAAC AACCTCTCGT CGTATCGTTC	14220

669

AGCTCTTGAA AGAAGCGGGT GCGACTGAGG TTCACGTTGC CATTGGAAGT CCTGCACTAG	14280
CGTATCCATG TTTCTACGGG ATTGATATCC AGACCCGTCA GGAGCTGATT GCAGCCAATC	14340
ATACGGTCGA AGAAACTCGC CAAATCATTG GTGCGGACAG TCTGACTTAT CTTTCAATTG	14400
ATGGCTTGAT TGAGTCGATT GGTATCGAAA CAGATGCGCC GAACGGTGGT CTCTGTGTCTG	14460
CTTACTTTGA CGGTGACTAC CCAACGCCTC TTTATGACTA CGAAGAAGAC TATCGTAGAA	14520
GTTTGGAAGA AAAGACCAGT TTTTACAAGT AGGCGACAGA TTCTCCATTA AAGAAAAGGA	14580
AAAAATAAAT GACAAATAAA AATGCATATG CCTCACGTCT CACTACTGAC TAAAGGCTTA	14640
AGCATTTAGT CAGTAGACGC TTTGTCTTAT AGGATCAAAG CTAGAGCCCT GACTAGTATT	14700
TTTAGATAAA AAGATGTTTT ATCTAAAAAT ACGTCGCAGT CTTTCTCAAA AAAAGAAAAG	14760
GAAAAATAAA ATGGCAAATA AAAATGCGTA CGCTCAATCT GGTGTGGATG TTGAAGCGGG	14820
TTATGAAGTT GTTGAACGGA TTA AAAAGCA CGTGGCCCGT ACGGAGCGTG CAGGTGTCAT	14880
GGGAGCTCTT GGTGGCTTTG GTGGTATGTT TGACCTTTCC AAGACTGGGG TTAAGAACC	14940
CGTCTTGATT TCAGGGACTG ACGGTGTCGG AACCAAGCTC ATGTTGGCTA TCAAGTACGA	15000
CAAGCACGAT ACCATCGGGC AGGACTGTGT GGCCATGTGT GTCAACGACA TCATTGCTGC	15060
AGGTGCGGAA CCCCTCTATT TTCTCGACTA CGTAGCGACA GGGAAGAATG AACCAGCTAA	15120
GCTAGAACAA GTGGTTGCTG GTGTGGCAGA AGGTGTGTG CAGGCTGGTG CTGCCCTCAT	15180
CGGTGGGGAA ACGGCTGAAA TGCCGGGCAT GTACGGCGAA GACCACTATG ACTTGGCTGG	15240
TTTTGCGGTC GGTGTGGCTG AAAAACTCA AATCATTGAC GGTTCAAAGG TGGTAGAGGG	15300
AGATGTTCTT CTCGGACTTG CTTCAAGTGG GATTCACTCA AATGGTTACT CTTTGGTTCTG	15360
TCGTGCTTTT GCGGATTACA CAGGTGAGGA AGTCCTACCA GAATTGGAAG GCAAGAAACT	15420
TAAGGAAGTT CTACTTGAGC CGACTCGTAT CTATGTCAAG GCTGTCTTGC CGCTCATCAA	15480
AGAAGAGTTG GTCAACGGCA TTGCCACAT CACAGGTGGT GGCTTTATCG AAAATGTCCC	15540
TCGTATGTTT GCAGATGACC TAGCTGCTGA AATTGATGAA AGTAAAGTTC CAGTGCTTCC	15600
AATTTTCAAA ACCCTTGAAA AATACGGTCA GATTAAACAC GAAGAAATGT TTGAAATCTT	15660
CAATATGGGT GTGGGACTTA TGTGCGCGT CAGCCCTGAA AATGTAGACC GTGTAAAAGA	15720
ATTGTTGGAT GAAGCAGTCT ATGAAATTGG TCGCATCGTC AAGAAAGAAA ACGAAAGTGT	15780
CATTATCAAA TGAAAAAAT AGCGGTTTTT GCCTCTGGTA ATGGCTCAAA TTTTCAGGTG	15840
ATTGCCGAAG AATTTCCAGT GGAGTTTGTG TTTTCAGACC ATCGTGATGC CTATGTGCTT	15900
GAGCGTGCAA AGCAGCTCGG CGTTCTGTCC TATGCTTTTG AACTCAAGGA GTTTGAGAGC	15960

670

AAGGCAGACT	ACGAAGCAGC	CCTTGTCGAA	CTCTTGGAAAG	AACACCAGAT	TGACTTGGTT	16020
TGCCTAGCAG	GCTACATGAA	AATCGTTGGA	CCAACCTTAT	TGTCGGCTTA	TGAAGGTCGG	16080
ATTGTCAACA	TTCATCCAGC	CTACTTGCCA	GAATTTCCAG	GAGCTCATGG	GATTGAGGAT	16140
GCTTGGAATG	CTGGCGTGGG	TCAGTCTGGT	GTGACCATTG	ACTGGGTGGA	TTCGGGTGTG	16200
GATACAGGCC	AGGTCATCAA	ACAGGTTCTG	GTGCCACGAC	TAGCTGATGA	TACCATTTGAC	16260
AGATTTGAAG	CTCGCATCCA	TGAAGCAGAG	TACAGGCTGT	ATCCGGAAGT	AGTGAAGGCT	16320
CTATTTACAG	ATTGACTTTT	TGATGATTCA	TATGATATCT	TTGATTTTAA	ATTGGAGTCA	16380
GTGTTTGTG	AAGACGGCTT	CAAACGGAGG	TATTTGTAAT	GTTAGAATCT	AAAAAAACAA	16440
CTCGATATGT	ATTTTATGTC	TATCTGATGT	TATTAACCTG	GGGAATCTTA	TTTAAGTTTG	16500
AAACAAATCC	TGAATTTATA	GCATTTTCTT	TAGCTCCAAG	GTATATCAAT	TGGATTCCAT	16560
TTTCAGAACC	ACTAATAGTC	GATGGAAAAA	TTGTTTTTGC	TGAAATGTTA	TTTAATCTGA	16620
TTTTCTTTAT	TCCATTAGGT	GTTTGTTCCT	CTTTGATAAA	AACTAATTTA	TCTAGTTTAA	16680
GAATAGTCGG	GACAGGTTTC	TTGATTAGTT	TATTGTTTGA	GTGCTTACAG	TATATTTTAG	16740
CAATAGGTAT	AACAGATATA	ACGGATTGTA	CTTTAAATAC	GCTAGGTGTC	TGTGTAGGCT	16800
TACTGATTTA	TCAAATTTTT	ATAAGAGTGT	TCAAATCACA	GACTAGAAAA	TGGATCAATA	16860
TCTTAGGTAT	GCTTAGCCTT	GGTTTTGCTT	ATCTTGTTTT	ACTGTTACTG	CATTTACTTA	16920
GTGTTTAACT	AATGATTAAA	AAGGAGAATA	TAATGACTAA	ACGCGTCTTA	ATCAGCGTCT	16980
CAGACAAAGC	GGGCATTGTT	GAATTTGCCC	AAGAACTCAA	AAAACCTGGT	TGGGAGATTA	17040
TCTCAACAGG	TGGAACATAAG	GTTGCCCTTG	ATAATGCTGG	GGTGGATACC	ATTGCTATCG	17100
ATGATGTGAC	TGGTTTCCCA	GAAATGATGG	ACGGTCGTGT	GAAGACCCCTC	CACCCAAATA	17160
TCCACGGAGG	GCTTCTCGCT	CGTCGTGACT	TGGATAGCCA	CTTGGAAGCG	GCTAAGGACA	17220
ACAAGATTGA	GCTCATTGAC	CTTGTGGTGG	TCAACCTTTA	CCCATTTAAG	GAAACTATCC	17280
TTAAACCAGA	TGTGACTTAT	GCTGATGCAG	TTGAAAATAT	CGATATTGGT	GGGCCATCTA	17340
TGCTTCGTTT	AGCAGCGAAA	AATCATGCCA	GTGTTACAGT	TGTGGTAGAT	CCTGCTGACT	17400
ACGCTGTGGT	TTTGGATGAA	TTGGCAGCAA	ACGGCGAAAC	CTCTTATGAA	ACTCGCCAAC	17460
GTTTAGCAGC	CAAAGTATTT	CGTCACACAG	CGGCTTATGA	CGCCTTGATT	GCAGAATACT	17520
TCACAGCTCA	AGTGGGTGAA	AGCAAGCCTG	AAAAACTCAC	TTTGACTTAT	GACCTCAAGC	17580
AACCAATGCG	TTACGGTGAG	AATCCTCAAC	AAGACGCGGA	CTTTTACCAG	AAAGCTTTGC	17640
CTACAGACTA	CTCCATTGCT	TCAGCCAAAC	AGCTCAACGG	GAAAGAATTG	TCATTTAATA	17700
ATATCCGTGA	TGCAGATGCT	GCTATCCGTA	TCATCCGTGA	CTTCAAAGAT	AGTCCAACCG	17760

671

TTGTGGCTCT CAAACACATG AATCCATGTG GAATTGGTCA AGCTGATGAC ATCGAGACTG	17820
CTTGGGACTA CGCTTATGAG TCTGACCCAG TGTCTATCTT TGGTGGGATT GTCGTCCTCA	17880
ACCGTGAGGT GGATGCTGCG ACAGCTGAGA AGATGCACGG CGTTTTCTC GAAATCATCA	17940
TTGCACCAAG CTATACGGAT GAAGCGCTAG CCATTTTGAT CAATAAAAAG AAAAATTGCT	18000
GTATCCTTGC CTGCCATTT AATGCTCAAG AGGCTAGCGA AGTGAAGCA GAATACACAG	18060
GTGTAGTCGG TGGACTTCTC GTGCAAAATC AAGACGTGGT CAAGGAAAGC CCAGCTGACT	18120
GGCAAGTGGT GACTAAACGT CAGCCAACCTG AGACAGAAGC GACTGCTCTT GAGTTCGCTT	18180
GGAAGGCTAT CAAGTACGTC AAATCAAATG GTATTATCGT GACCAACGAC CACATGACAC	18240
TTGGTGTGG TCCAGGTCAA ACCAACCGTG TGGCTTCTGT TCGCCTTGCC ATTGACCAAG	18300
CCAAAGATCG TCTGGACGGG GCGGTCCTTG CTTGAGATGC CTTCTTCCCA TTTGCGGATA	18360
ACGTGAAGA AATCGCCAAA GCAGGAATTA AGGCCATCAT CCAGCCCGGT GGCTCTGTCC	18420
GTGACCAAGA ATCCATCGAA GCAGCGGATA AATACGGCTT GACTATGGTC TTTACAGGTG	18480
TGAGACATTT TAGACATTAA GAAGATAAAA GGAAGAAAA CAGTTTCTTT CCTTTTGG	18540
CTTAAATAC TAACTGAAAC AAGATTAAAA CGAACTTTTT TGATATAATG TTGGTAAATA	18600
ATTCGCAAAA GAGGTGAGG AATGAACTG CTTGTTGTCG GTTCTGGTGG TCGTGAGCAT	18660
CGGATTGCTA AAAAGTTACT TGAATCAAAA GACGTGGAAA AAGTCTTTGT AGCTCCTGGG	18720
AATGATGGA TGAATCTGGA TGGTTTGGAA TTGGTAAATA TCTCTATTTT CGAACATTAT	18780
AAATTGATTG ACTTCGCAAA GACCAATGAT GTTGCTTGA CTTTATCGG TCCAGATGAT	18840
CCCCTTGCTG CTGGTATCGT GGATGATTTT AACCAAGCTG GACTTAAGGC CTTTGGTCCG	18900
ACTAGGGCTG CAGCGGAGCT GGAGTGGTCC AAGGATTTCC CCAAGGAAA CATCGTCAAA	18960
TACGGCGTTC CGACAGCAAC ATATGGCACA TTTTCAGATT TCGAGGAAGC CAAAGCCTAT	19020
ATCGAAAAGC ATGGTGCGCC TATCGTAGTC AAGGCGGATG GCTTGGCACT TGGGAAGGGT	19080
GTCGTCGTTG CTGAGACGGT TGAGCAACCG GTCGAAGCCG CTCATGAGAT GCTTTTGGAC	19140
AATAAATTG GTGACTCAGG TCGCGCGCTG GTTATTGAGG AATTCCTTGA AGGAGAGGAA	19200
TTTTCACTCT TTGCCTTTGT CAATGGTGAT AAGTTCTACA TCATGCCAAC GGCTCAGGAC	19260
CACAAACGTG CCTATGATGG CGACAAAGGG CCTAACACGG GTGCTATGGG TGCCTATGCG	19320
CCAGTCCAC ACTTACCACA GAGTGTAGTT GATACAGCGG TTGACACCAT TGTCAGGCCA	19380
GTTCTAGAAG GGGTGATTAA AGAAGGTCGC CCTTATCTGG GAGTTCTTTA CGCAGGGCTT	19440
ATCCTGACAG CTGATGGACC GAAAGTCATT GAGTTCAACG CTCGGTTCGG AGATCCAGAA	19500

672

ACTCAGATTA TCTTGCCCTCG CTTGACCTCT GACTTTGCTC AAAATATCAC AGATATCCTG	19560
GATAGCAAGG AGCCAAATAT CATGTGGACG GACAAGGGTG TGA CTCTGGG TGTGGTTGTC	19620
GCATCCAAGG GCTACCCGCT AGACTATGAA AGGGGCGTTG AGTTGCCAGC CAAGACAGAA	19680
GGCGATGTCA TCACCTACTA TGCAGGGGCT AAGTTTGCGG AAAATAGCAG AGCACTGCTC	19740
TCAAACGGCG GACGAGTTTA TATGCTCGTT ACCACAGCAG ATACCGTCAA AGAAGCCCAA	19800
GCCAGCATAT ACCAAGA ACT ATACCAACAA AAAATAGAAG GACTCTTCTA CCGAACAGAT	19860
ATCGGAAGCA AGGCAATTAA GTAAAGATAT AAGAATAACG CGCCGTAGTC GCCAAACACG	19920
ATAATGGTCG TCGTGGTGAA AAGACCAGAA CAGTGAATGT TCTGGTCAGG GGGAAACTTG	19980
GAGACCTTAG GCTCAAAGTT TAGGAATGAA ACCGAAGGTT TGCTTCCGCC TCCATCACCT	20040
AAGACCATTA TCAAAAAGAA AAATAAAAAT TCACAAAATA CGTTAATGAT CGTATGGTTT	20100
GCGAGCGTTA GCGAGCTAAT ATAGAACAAT CACCGCCGTT GTGAAAGAAC GATTGGATGA	20160
TAATCCAATC GTTCAGGGAA ATTGGAAGAC CTTGGGTTTC CAATTTAGGC ATGAGACACC	20220
TTTGGTGGCT GCTGCCGTCC CTCACAAGCT AAGGTGATTG TTGAAAAAGA GGAAAAAGGA	20280
GAAGAAATGA AACCAGTAAT TTCCATCATC ATGGGCTCAA AATCCGACTG GGCAACCATG	20340
CAAAAAACAG CAGAAGTCCT AGACCGCTTC GGTGTAGCCT ACGAAAAGAA AGTTGTTTCC	20400
GCACACCGTA CACCAGACCT CATGTTCAAA CATGCAGAAG AAGCCCGTAG TCGTGGCATC	20460
AAGATCATCA TCGCAGGTGC TGGTGGCGCA GCGCATTTGC CAGGCATGGT AGCTGCCAAA	20520
ACAACCCTTC CAGTCATTGG TGTGCCAGTC AAGTCTCGTG CTCTTAGTGG AGTGGATTCA	20580
CTCTATTCTA TCGTTCAGAT GCCGGGTGGG GTGCCTGTTG CGACCATGGC TATCGGTGAA	20640
GCTGGAGCGA CTAACGCAGC TCTCTTTGCC CTCGCTCTCC TCTCTGTAGA AGATAAGTCC	20700
ATTGCCGATG CACTTGCCAA CTTTGCTGAA GAACAAGGAA AAATCGCAGA GGAGTCGTCA	20760
AATGAGCTCA TCTAAAACAA TCGGAATTAT CGGTGGCGGT CAACTGGGTC AGATGATGGC	20820
CATTTCTGCT ATCTACATGG GCCACAAGGT TATCGCGCTG GATCCTGCGG CGGATTGCCC	20880
GGTCTCTCGT GTGGCGGAAA TCATTGTGGC ACCTTATAAC GATGTAGACG CCCTCCGTCA	20940
GTTGGCAGAC CGTTGCGATG TCCTCACTTA TGAGTTTGAA AATGTCGACG CTGACGGTTT	21000
GGATGCCGTT ATCAAGGATG GACAACTCCC TCAAGGAACA GATCTGCTCC GCATTTGCGA	21060
AAATCGTATT TTTGAAAAGG ACTTTTGTGTC AAACAAGGCT CAAGTCACTG TGGCACCCCTA	21120
CAAGGTCGTG ACTTCTAGCC TAGACTTGGC AGATATCGAC TTGTCGAAAA ACTATGTCCT	21180
CAAGACTGCG ACTGGTGGCT ACGATGGTCA TGGACAAAAG GTTATTCGTT CAGAAGCAGA	21240
CTTGGAAGCA GCCTATGCGC TAGCAGACTC AGCAGACTGC GTCTTGGAAG AATTGTCAA	21300

CTTTGACCTT GAGATTTCTG TCATCGTGTC AGGAAATGGC AAGGAGGTGA CGTTTTTCCC	21360
AGTTCAGGAA AATATCCACC GCAACAATAT CCTGTCTAAG ACCATCGTAC CAGCCCGCAT	21420
TTCTGAAAGT CTAGTAGACA AGGCTAAAGC TATGGCAGTG CGAATCGCAG AACAACTCAA	21480
CTTGTCTGGA ACTCTCTGTG TGGAAATGTT TGGCAGAGCT GATGACATCA TTGTCAATGA	21540
AATCGCCCCA CGACCACATA ACTCTGGGCA CTATTCTATT GAAGCCTGTG ATTTCTCTCA	21600
GTTTGACACC CATATTCTGG GTGTTCTCGG AGCACCATTA CCAGTCATCA AACTCCATGC	21660
GCCAGCCGTT ATGCTTAATG TCCTCGGTCA GCATGTCGAG GCTGCTGAAA AATATGTCAC	21720
AGAAAATCCA AGCGCCCACC TCCACATGTA TGGTAAAATA GAAGCAAAGC ATAATCGTAA	21780
GATGGGACAT GTGACTTTGT TTAGTGATGT GCCGGATAGT GTGGAAGAGT TTGGGGAAGG	21840
GATTGATTTT TAGGACAAGT CTATGATACA AATTATCGTT AATACATTTA TTGAAAAGTA	21900
TAAGACTGGA GCAGTTGTTG AAGTGTGTA TGCCAGTGCT GACCAAGATA AGGTACAAGC	21960
TAAATATGAA GAACTAGCTG CACAATACCC CGAAAATTAT TTAGCTATCT ATAATGTACC	22020
GCTGGATACG GATTTGAATA CACTAGATCA TTACCCGTCT GTGTTTATTG GAAAAGAGGA	22080
GTTTGAGTAG AAATCTTGGT TTACCTAGAT AGCTTATTC CAACAGCTTA AGAAGAAAGG	22140
AAAAATTAAC ACATGATCAA CCGTTACTCT CGCCCTGAGA TGGCGAATAT TTGGAGTGAA	22200
GAAAATAAAT ACCGTGCTTG GCTTGAGGTG GAAATCCTCT CTGACGAGGC ATGGGCTGAG	22260
TTGGGGGAAA TCCCTAAGGA AGATGTGGCT TTGATTGCGA AGAAGGCGGA CTTTGACATC	22320
GACCGTATTT TGGAAATTGA GCAGGAGACG CCCCACGATG TGGTGGCTTT CACGCGTGCG	22380
CTTCTGAGA CTCTTGGTGA AGAGCGCAAG TGGGTCACT ATGGGTAAAC TTCTACTGAC	22440
GTGGTGGATA CTGCTTATGG TTACCTCTAC AAGCAGGCCA ACGACATCAT CCGTCGTGAC	22500
CTTGAAAAC TCACTAATAT CATCGCTGAC AAGGCCAAGG AGCACAAGT CACCATCATC	22560
ATGGGGCGTA CTCATGGTGT GCACGCTGAG CCGACAACCT TTGGTCTTAA ATTAGCAACT	22620
TGGTACAGCG AAATGAAACG CAATATCGAG CGCTTCGAGC ATGCGGCTGC TGGTGTAGAA	22680
GCTGGTAAGA TTTCTGGTGC GGTGGAAC TTTGCCAATA TCCCACCATT TGTAGAGGAG	22740
TATGTCTGCG ATAAACTTGG CATCCGTGCC CAAGAAATCT CTACACAAGT CCTTCCTCGT	22800
GACCTTCACG CTGAGTACTT TGCGGTCTT GCCAGCATTG CGACTTCAAT CGAACGTATG	22860
GCGACTGAGA TTCGTGGTCT ACAAAAATCT GAGCAACGCG AAGTAGAAGA GTTCTTTGCT	22920
AAAGGGCAAA AAGGGTCTTC AGCAATGCCT CACAAACGCA ACCCAATCGG TCTGAAAAT	22980
ATGACTGGTC TGGCGCGTGT CATTCGTGGT CACATGATTA CGGCTTATGA AAACGTCGCT	23040

674

CTCTGGCATG AACGCGATAT TTCTCACTCA TCAGCTGAGC GTATCATCAC ACCAGATACG	23100
ACCATTTTGA TTGACTACAT GCTCAACCGT TTTGGAAATA TCGTCAAGAA CTTGACAGTC	23160
TTCCCAGAAA ATATGATCCG AAACATGAAC TCGACTTTTG GTCTTATCTT TAGCCAACGG	23220
GCTATGTTGA CATTGATTGA AAAAGGCATG ACCCGTGAGC AAGCCTATGA CTTGGTGCAA	23280
CAAAAACAGC CTACTCTTGG GACAACCAAG TAGACTTTAA ACCACTTCTT GAGGCAGATT	23340
CAGAAGTAAC ATCAGCTCTC ACACAAGAAG AAATCGATGA AATCTTCAAC CCAGTTTATT	23400
ACACCAAACG AGTGGATGAT ATCTTTGAAC GTCTTGGACT AGGTGATTAA TTA AAAAATA	23460
AACAGCGAGC TTCAATCTCG CTGTTTATTT TTTATCGAAA AGACTTAGTC TTCTTTTCTT	23520
TTAGTGAGTC CATAGGCTGC TAGTGTGGAC ATGAGTCCTG CGACTACTAG TCCTGCAGAA	23580
TCGTGAGTTC CTGTTTCAGG AAGTTTTTTC TCTGTTACCA CAGGAGCTGG ATCTTGAGGA	23640
AGAACTTTGC TTTCTCAGC AGGAGCAGTT GATGGAGCTG GTTGGCTTGG GATTTCTAGT	23700
TTTGGTTTTT CTTCAGCAAT AGCGGCTTGT CCGTTTTCAT CGCCTACATG TGTTACCATA	23760
GTTCCGACTT CGACTATTTG AGTAACGGCT TCCTGTGCTA CGACACTATT TACAAGTGTT	23820
TTCACTTCCT TACCATCGGC AGAAGTGCTC ACAGAGTAGA AGTTGCTACG ATGTCCATTG	23880
ACGCCCTTAG TAATGACTTG TGTTTTCCT TTGAGTAAGA GTGGATTTTC ACAAGTCACT	23940
GTGGTAAATG GAATTTCTTC TTCTTGGATA TCCAGTCTAG GTTTTACCTC AGTAGTTGGT	24000
GCAAGACCAC TTTATCACC CTTGTGAGTT ACAGGAGCGC CAACTTCAAC CACTTGGTTT	24060
ATAACTTCTT TGGTTACCTG GCTATCAAGG ACTGTTTCTG TTGTTTTTCC ATTTTCAGTG	24120
AGTACAGAGA TGTAATGAGT TCGTTCACCT TTGACTCCTG CTGTGATAAT ATTTTCCTGA	24180
CCGGCTGGGA GGTTAGGATT TTCTTTCTTG ATAACITCAA ATGGAATTTT TTAGTTCTT	24240
GTGATGAGTT CTGGTCTGGT TTCAACATTG GCAGCCACTT CATTTTCATC TAGGCTTCCT	24300
GAATGAGTTA CAGCTGGTTT GAGGCCTTGA AGAGCGGCTT TTAGGTTGGC TACAAGCGTG	24360
TCAAGCTCAG CTTGTTTATT ACGGTTGAGG TTGTAATTTA GAGCTGTTTT AGCTGCGTCA	24420
AGGGCCTCAA GACTTTCTTT ACTATATCCT TCTAAGTTTG TAGGAATTTT AGCTAATTCT	24480
TCGCGGAGAG CATTATAATT AGCACGAAAG TAGTCTTTGT TGTGGTCTGC AAAGGCAGTC	24540
ATGAGTTCAA AGATTTCTCTC TTCCTTGTAT TCAGCGCTTG GTCTATCTGC CCAGATTGAA	24600
AGCATACTTC CGACTGTTGG AAGATCTACT TCAGGATATT TGGTAGAAGC TAGTTGATTG	24660
AATGGTGGTT TTCCAGTATT CTCAATAGCT TTCTTGAGGA AACCACCACC ATCTTCTGGT	24720
TTTTGACCAA GAATGTAGTA CCAGTCACCG TTGGTATTCA AGAATTTATA GCCTTTGCTT	24780
GCTAGGTATT GAGGTGATGC GAGGTTATAT CCCCACCAGC CTTTAGACCA GTAAGAAATC	24840



675

AAGACATCTT TGTCAAAC TG AACATCGTCC TTGTCTTCAT AGTAGAAGCC ATCGTTGAAG	24900
GCCATTGGTT GAAGCCCTCT TTCTTTGGCC ATAGCTGCCA GGGTGTGGC ATATTCGGCA	24960
AATTTGCCAT AGAGTTGATA CCACTTGAGG TAGTACCAGC CTTGGGCACT AGTCGCATCG	25020
TTGGCGTATT CGTCAGTACC AAAGTTGAAA ATCTTTGTTT TACCTGCAAA GAAGTCCATG	25080
TATTTACCGA TGAGGGCTTT TACAAAGTTC ATCGCTTCTT CGTTTTTCAA GTCCATAGTT	25140
GT TTTTGAAA CTTTATCAAA GTGGGCTTGA GGATTTTAA TACCTAATTT TTCCATGGCA	25200
ACCAGCATAG CATCCATGTG ACCTGGACTG TTAATAGCTG GGATGAGACC GATGTCCTTA	25260
GATTTAGCGT ATTCAATTAG CTCTGTTACT TCTGCCTGTG TTAGTGCACT ACCGTTTGGA	25320
TCGTCGTAGT AAGCTTTAGT TCCTTCGATA ATAGCTTTTT TAACGTCATC ACTAGCATAG	25380
GT TTTTCCGT TGGCAGTAAT GGTCAATCA TCGAGTAGAA AGCGAAGTCC GTCATTTCTT	25440
AGAAGGAGAT GGACATCAGA ATATCCGAGC TCACTGGCCT TGTCTACGAT GCGTTTGAGC	25500
TGGTTCAGAG TAAAGTATTT GCGTCCAGCA TCGATTGAGA TTACCTTGTT TTTGGCAAGT	25560
TTTTCAACCT CACGTTTAGC TTCTTCTTCT TTTTGAGCTT CAGGCGTGAG GGTCAAGTTG	25620
TTGACAGTTT CTTGAAGTTT AGCAATGGCT TGATCAATCG TATCTTGTTG GGCACGGCTA	25680
AGGTTGCTAT CGAGAGAGCG AATAGCTTTT TCAGCTTCTT TTACGGCCGT GACGCTTTCT	25740
GCACTATAAC GGTTCAGGTC TTTTGGTACC TCGTTAAGTG CTTGCTCTGC AGATTCATAA	25800
TCAGCTGCCA AGTATTCAGC GTTGGCATTT GCAAAATGAC GCATGAGTTT GAAGAGGCGT	25860
GATGGTGAAT AACGTGCAGA TGGAGTGTCA GCCCAAGCAG CTACCATACC ACCGATGATT	25920
GGGATATCAG CTCCTTCTGT TTTTGGTACA GAAGTGATTG GTGTGTTTTT AATACCATTG	25980
AGCCCCGTAT CGAGATTGTA CCAGCCTTGG CCATCAGCGT TTCGTCCAAG AACGTAGTAC	26040
CAAGCATCAT TGGTATTAAG GATTTGGTGA CCTTTTTCAG CTAGTAGTTT AGAAGAAGCG	26100
ACATCGTAGC CTCCCCAACC ACCAGTCCAC ATAGAAACGA TGATGTCCTT GTCAAACTA	26160
CCAAAGCTTG TGTGCTATT GTAGTAGATA CCGTCGTTAA AAGCCATTGG TTTGAGACCG	26220
TGCGATTTTA CAATACGAGC GAGGTCAATTG GCGTAGGCAA TAAATTTTTC ATAGCCTTTT	26280
ACAGGGTAGC CTTGCTTTGG ATAGTATTTA TCAGCTTGAA GCACACTCCA ACCTTTAGCA	26340
TCTGTGCGAT CATTGGCATA TTCATCAAGT CCGATGTTGA AGATTTCAGT CTTTTTCGCG	26400
AAATAAGCAG CATACTTGTC GATAAGGGCT TTTGTAAAAG CGACAGCTTG TTCGTTGTCA	26460
AGATCGACAG TACGGGCTGA TTTCTTCCCA AAATAGCTAA AGTTAGGCTT TTGGATTCCC	26520
AATCTTTTCA TGGCATTGAG AATCGCATCC ATGTGTCCAG GACTATTTAC TGTCCGAATG	26580

676

AGACCGATAC CTTTATCTTT GGCATAGTTA ATCAGATCTG TCATTTGACT TTCTGTAAAG	26640
TGATTGCCGT TTGGATCGTT GTAATAATCA TTTGTACCTT TTTCAATGGC GCGTTTGACA	26700
TCGTCACTGG CATAGGTCTT GCCGTTAGCT GTGATGCTCA TATCGTCCAA CATGAAACGG	26760
AGTCCATCAT TTCCGACTAA TAGGTGTAAA TCAGTGTAGC CATAATGTTT CGCTTTATCG	26820
ATGATTTCTT TGAGCTGTTT TGGTGAGAAA TATTTACGTC CAGCATCAAT AGAAACAATT	26880
TTCTTTTTCG CTAGTTTTTC ATTTACAGTT GCAGCACGTT CCTTTCCTGC CTCTGTGCGC	26940
GGTTTGTGAG CCTCTGCTTT CGCTTCATCT TTTTATAGCTG GTTTATCCTT GTCAGTCTTG	27000
TCTGTATTTG ACTCTTTAGA ATCAACCTCT TTCGCTTCTT CCTTTTTAGG GCTAGCTTCT	27060
TCTGCCTTTT TATTAGCAGT TTCTTTTCA GCAGAAGTTG GAGTTACCAC TTCTGCTTTA	27120
TCACTAGGAG TTGAACATAA TTCCTCTTGT GGTTTTTCTT CTGTTTTTGG AAGACTAGCT	27180
ACCTTATCAG TAGCTGGAGT TTCTGTTTCT ACAGTTTTTG GAGCTTCTGG TTGAAGCACT	27240
GCTTTAGGTG TTTCTCAGT CCGATTTTCG GATGATTGAG GGAATCAGA AACCGTATGG	27300
ATGGTCGGTT GGTTTTCTGT AGTAGTAGGA GTAACCTCAT CGGCTGCAAC AGTCTGTGCT	27360
TGGAAGGCAA ATCCAATTAG AACAGAAGCT GCTCTACAG CGTATTTACG AATAGAAAAA	27420
CGCTGTGTTT TTTCATGTTT CATTGCAAAA CCTCCTGATT GCATTGTTAT ATTGATAGCG	27480
ATTATATAAA TCAACGCCTT TATTTTATTT CTTATATTAA TTTCTTATAT TAACGAGAGT	27540
CAAGAGGAGA TGACAAAAAA CTATAATAAG TATAAAAAAA TATAAAATTT AAACCTAAGA	27600
TTTCAGATTG GTCGGAAAAA ATACGTATAT ATATCTAGTA TAATTTTTGG TTCTATTTCT	27660
ATAAAATATT CCACAAATTA TAGAATTTTC CAAAAATAGG TAAGCGCTAC CTTTTGGTG	27720
TAGTATAATA AGCATAGAAA AAGCCCAAGC GATTAGCTCA GGTTTTCTTC TTAGTGATCA	27780
CGGTCACATG AGATAAATTT AATCTTGTAG TAATCAGATC GTTTGTAAGT TTCACTGTAT	27840
TCTAAACTT GGCCAGTTGA TTCGAGTTG GTGATTTTAG TTTGTAGGAC AGTAGGGAAT	27900
TGTTTCATCGA CTCCGAGGAC TGAAGCTGCA TGTTCGGAG TTGGAAAGAC TATTTCTTG	27960
ATTTCTTCAA AGTGTTTCATC ATTCATGTGA ATGTGGTAGT CTAACCTGAA ACGATTATAG	28020
ATAGAACTAT AGTATTCAAG GTTTGGATAA TTTGCGTTGA TATATTGTTT TGGGATGTAG	28080
GATGTATGGT AGATATAAAC GACACCGTTT GATTGCGGGA TACGTTCAAT CTTGTAGTAG	28140
AATTGATCGC CGCGTAGACC CAATTTTCC AAGTAAACAA GCTTGTTC CCGTTCAATT	28200
GAAAGAACAG TTACCTTATC ATCTTTAGCA TTGAAGAGTT CAATATCTGA AAACCTTACA	28260
AGCTTGTGTT TGCGTGACG TGAAACGAAG GTTCCTTTTC CTTGTTGGCG GACAATATAG	28320
CCATCTTTGG CAAGGTCGTT TAAGGCGCGA ACAACTGTGA TAGAGCTGAC ATCGTACATT	28380

677

GAAATGAGTT CTGCTTCAGT GTAAAATTTA TCTECACTGC TAAACTGCCC AGAGATGATT 28440

TTATTTTTTA ATTCTCTTT TATGTATTGA TGG 28473

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6749 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGATGGGT GGTATGCGAG GATACAGTTC TGAAAATCCC CGTTACTTAA TTAATGGACG	60
CGAAGTCACA CCTGAGGAAT TTGCTCACTA TCGTGCGACT GGTCAATTAC CAGGAAATGC	120
AGAAACTGAT GTGCAAATGC CACAACAGGC ATCAGGTATG AAACAAGGCG GTGTCCTTGC	180
AAAAGTAGGT CGAAACTTAA CAGCAGAAGC GCGTGAGGGC AAGTTGGATC CTGTTATCGG	240
ACGAAACAAG GAAATTCAAG AAACATCTGA AATCCTCTCA CGCCGCACCA AGAACAATCC	300
TGTTTTGGTC GGAGATGCAG GTGTTGGTAA GACAGCAGTT GTCGAAGGTC TAGCGCAAGC	360
CATTGTGAAC GGAGATGTTT CTGCTGCTAT CAAGAACAAG GAAATTATTT CTATTGATAT	420
CTCAGGTCTT GAGGCTGGTA CTCAATACCG TGGTAGCTTT GAAGAAAATG TCCAAAACCT	480
AGTCAATGAA GTGAAAGAAG CAGGGAATAT TATCCTCTTC TTTGATGAAA TTCACCAAAT	540
TCTTGGTGCT GGTAGCACTG GTGGAGACAG TGGTTCTAAA GCACTTGCCG ATATTCTCAA	600
GCCAGCTCTC TCTCGTGGAG AATTGACAGT GATTGGGGCA ACAACTCAAG ACGAATACCG	660
TAACACCATC TTGAAGAATG CTGCTCTTGC TCGTCGTTTC AACGAAGTGA AGGTCAATGC	720
TCCTTCGGCA GAGAATACTT TTAATAATTCT TCAAGGAATT CGTGACCTCT ATCAACAACA	780
CCACAATGTC ATCTTGCCAG ACGAAGTCTT GAAAGCAGCG GTGGATTATT CTGTTCAATA	840
CATTCTCAA CGTAGCTTGC CAGATAAGGC TATTGACCTT GTCGATGTAA CGGCTGCTCA	900
CTTGGCGGCT CAACATCCAG TAACAGATGT GCATGCTGTT GAACGAGAAA TCGAAACGGA	960
AAAAGACAAG CAAGAAAAAG CAGTTGAAGC AGAAGATTTT GAAGCAGCTC TAACTATAA	1020
AACACGCATT GCAGAATTGG AAAGGAAAAT CGAAAACCAC ACAGAAGATA TGAAAGTGAC	1080
TGCAAGTGTC AACGATGTGG CTGAATCTGT GGAACGAATG ACAGGTATCC CAGTATCGCA	1140
AATGGAAGCT TCAGATATCG AACGTTTGAA AGATATGGCT CATCGCTTGC AAGACAAGGT	1200
GATTGGTCAA GATAAGGCCG TAGAAGTTGT AGCTCGTGCT ATCCGTCGTA ACCGTGCTGG	1260

678

TTTTGATGAA GGAAATCGCC CAATCGGCAA CTTCTCTTT GTAGGGTCTA CTGGGGTTGG	1320
TAAGACGGAG CTTGCTAAGC AATTGGCACT CGATATGTTT GGAACCCAGG ATGCGATTAT	1380
CCGTTTAGAT ATGTCTGAAT ACAGTGACCG CACAGCTGTT TCTAAGCTAA TTGGTACAAC	1440
AGCAGGCTAT GTGGGTTATG ATGACAATAG CAATACCTTA ACAGAACGTG TTCGTCGCAA	1500
TCCATACTCT ATCATTTCTCT TGGATGAAAT TGAAAAGGCT GACCCCTCAAG TTATTACCCT	1560
TCTCTCCAA GTTCTAGATG ATGGTCGTTT GACAGATGGT CAAGGAAATA CAGTAAACTT	1620
CAAGAACACT GTCATTATTG CGACCTCAAA TGCTGGATT TGGCTATGAAG CCAACTTGAC	1680
AGAAGATGCG GATAAACCAG AATTGATGGA CCGTTTGAAA CCCTTCTTCC GTCCAGAATT	1740
CCTCAACCGC TTTAATGCAG TCATCGAGTT CTCACACTTG ACTAAGGAAG ACCTTTCTAA	1800
GATTGTAGAT TTGATGTTGG CTGAAGTTAA CCAAACCTTG GCTAAGAAAG ACATTGACTT	1860
GGTAGTCAGT CAAGCGGCTA AAGATTATAT CACAGAAGAA GGTACGACG AAGTCATGGG	1920
GGTTCGTCCT CTCCGTCGCG TGGTTGAACA AGAAATTCGT GATAAGGTGA CAGACTTCCA	1980
CTTGGATCAT TTAGATGCTA AACATCTGGA AGCAGATATG GAAGATGGCG TTTTGGTTAT	2040
TCGTGAGAAA GTCTAAGACA GAATTTTGAG GATAAAAAAG AAGGAGCCAG CTGAAAAAAA	2100
CTGGTTCCTT TTTAGGTACG ACAGGCATGT CGTATAGTAG AAGTGTATTA TTCTAGTTTC	2160
AATATACTAT AGTAGCTCAG AAGTCGGTAC TTAACGTGC TATATCAAAA CCAGTCCTGG	2220
AAAAACGTGG ACTGGTTTCG TGTTTGGATT ATTACCTTGA ACGACATGCG TTAAGTTA	2280
GTTGAACCGC CGTATGCCGA ATGGTACGTA CGGTGGTGTG AGAGGGGCTA GAGATTATCC	2340
CCTACTCGAT TTAAATCAC ATGACGTTCA AAGGCATCAT CTGAAATCCC TTGTTCCAAG	2400
ATGAGTTTTC CCCATTCTTT AGCAGAGAAG AGGCTGTGGT CTTGTAGTT TCCGCAAGAT	2460
TCGATGGTTG TCCCTGGGAC ATCTTCCCAA GTAGTAGTTT CAGCGATTTT CTTGAGCGAA	2520
TCCTTGATAA CAGCTGCGAT TTTAGCACTG GTGTGACGTC CCCACATAAT CATGTGGAAG	2580
CCTGTGCGGC AACCAAATGG TGAACAGTCA ATCATGCCGT CAATGCGGGT ACGGATGAGT	2640
TTGGCTAAGA GGTGCTCGAT AGTGTGAAG CCGGCAGTAG GGATAGAGTC TTCGTTTGGT	2700
TGCACCAAGC GAATATCATA ATTGGAGATG ATGTCTCCTT TTGGTCCTGT TTCTTCCCA	2760
ATCAAGCGAA CATAGGGTGC TTTGACAATG GTGTGGTCAA GTTCAAACT TTCGACAATA	2820
ACTTCTTTTG ACATGGTAAA TCCTTTCAGT TTTCTTCTCT CATTATATCA TAAAGGTTGC	2880
TCCTGAGACA GAGAGAAAAC CTCTCCGAGG CTGGAGAGGT TGAAATCTTT ACTTACGATA	2940
TAAGCGGTCG TATTGGTAGT ATGGGTCAAA GGTACGTTG ATACCCAGTT TACGAAGGAC	3000
ATTCTTGTCT TCATCAGTCA AGATGATGGT TGAGTGGGCT TCGTTCCTT TGAGGTTGCC	3060

679

GAGTTCTTCC ATAGCGCGGG CAGCATCAGG ATTTTCTGTA GCTGTGATAG CAAGTGCAAT	3120
CAGGATTTC A TTTGAATGAA GCGGTGGATT GCGGCTACCG AGATGATCGA TTTAAGACC	3180
TTGGATTGGC TTAACAACTT CAGGCTCGAT TAGTTTTACT TCTTTAGCGA TGTCAGCTGA	3240
TTTTTTGATG GCGTTGATCA AGGCAGCGGC TGTAGGACCA AAGAGTCTG ACTTCTTACC	3300
AGTGATGATT TCCCCATTTG GCAATTCAAA GGCTAGGGCT GGTCCACCAG TTTCTTCTGC	3360
TTTTTTGGCGC GCAACGACAG CAACCTTACG GTCTGCAGGT GTGATACCGA GGTCGTTTAT	3420
GAGCAACTCA ATTTTCTTGA CGGCAGCTTC GCCAACTTT TCAGCTTTGA AGTCAAGAAC	3480
TGTTTGATAG TAACGGCGGA TGATTTCTTG TTAGAAGCT TCGACAGCGG CCTCGTCATC	3540
TGTAATAGCG AAACCAACCA GTTGACACC CATATCTGTC GGTGAAGCGT ATGGTGATTT	3600
TCCGAGAATA CGTTCCAACA TCGGTTTGAG CACTGGGAAG ATTTGATAT CACGGTTGTA	3660
GTGACAGTG GTTCTCCAT AGGTTTGAAG ATGGAAGGGG TCAATCATGT TGACATCATC	3720
AAGGTCAGCT GTGGCAGCTT CATAAGCCAA GTTAACTGGA TGATGAAGGG GAAGATTCCA	3780
AACAGGGAAG GTTCAAATT TAGCGTAGCC AGATTTGATG CCATTGATTT GGTCGTGGTA	3840
CATATTGGAC ATACACGTTG CCAATTTTCC AGAACCAGGT CCAGGAGCGG TTACGACAAT	3900
CAAGTTGCGA CTGGTTTTGA TGTAGTCGTT TTTGCCCATG CCTTCTGGGG AAATGATGTG	3960
ATCCATATCC GTCGGATATC CTTTGATTGG ATAATGAAGA TAAGAATCAA TTCCGTTTTT	4020
CTCAAGTTGA TTGCGGAAGG CATCTGCACC GGGTTGGCCA GCGTATTGTG TAATGACAAC	4080
GGAACCAACA AAAATCCCTA ATTCATTGAA TTTATCAATC AAACGAAGAA CTTCTTGGTC	4140
ATAAGAAATG CCTAAGTCGC CACGTGCTTT GGAATGTTCA ATGTTGCTAG CATTAATGGC	4200
AATCACAACC TCAACCTGCT CTTTCAATTC TTGCAAGAGC TTGATTTTGT TGTCAGGTTT	4260
ATAACCAGGA AGGACACGAG CAGCGTGGAA ATCTTCTAAC ATTTTACCGC CAAACTCTAA	4320
GTAGAGCTTG CCGTCAAATT GGTAAATGCG CTCCAAAATA TGGTCGCGTT GTAAATTCAA	4380
ATATTGTTCA GAACTAAAAG CTGTTTTTTT CATTTTTTTA CCTCTGGACT CTATTATAAT	4440
AAAAAATTGG AAGTTAGGAA ACTACGGAGC TAAAAAGAA ATTAAAAAGA TTAAGCAAAC	4500
GCTTGACAAA AATTTTAAAA AGTGCTATCA TAGACTATAG ATTATGAAAA TAATGAGGTA	4560
AACAGATGCA AGAAAAATGG TGGCACAATG CCGTAGTCTA TCAAGTCTAT CCAAAGAGTT	4620
TTATGGATAG TAATGGAGAT GGAGTTGGTG ATTTGCCAGG TATTACCACT AAGTTGGACT	4680
ATCTAGCTAA GCTAGGAATC ACAGCAATTT GGCTTTCTCC CGTTTATGAC AGCCCTATGG	4740
ATGATAATGG CTATGATATT GCTGATTATC AAGCGATTGC GGCTATTTTT GGAACCATGG	4800

680

AGGACATGGA TCAGCTGATT GCAGAAGCTA AGAAGCGTGA CATTCTGTATC ATCATGGACT	4860
TGGTGGTCAA TCATACCTCA GATGAACATG CTGGTTTGT CGAAGCCTGT GAAAATACTG	4920
ACAGCCCTGA GCGAGACTAC TATATCTGGC GCGATGAACC CAATGACCTA GATTCTATCT	4980
TTAGTGGGTC TGCTTGGGAA TACGATGAAA AGTCAGGTCA ATACTATCTC CACTTTTCA	5040
GCAAGAAACA GCCGGATCTC AACTGGGAAA ATGAAAAACT TCGCCAGAAA ATTTATGAGA	5100
TGATGAACTT CTGGATTGAT AAAGGTATTG GTGGTTTCCG TATGGATGTT ATTGACATGA	5160
TTGGCAAAAT TCCTGACGAG AAGGTAGTCA ATAATGGTCC TATGCTCCAT CCCTATCTCA	5220
AGGAAATGAA TCAGGCGACC TTTGGAGATA AGGATCTCTT GACAGTAGGG GAGACTTGGG	5280
GAGCAACTCC AGAGATTGCC AAGTTCTACT CTGATCCAAA GGGCCAAGAA TTGTCTATGG	5340
TCTTCCAGTT TGAACATATC GGTCTTCAGT ATCAGGAAGG TCAGCCTAAA TGGCACTATC	5400
AAAAAGAGCT GAATATCGCT AAGTTAAAAG AAATCTTCAA CAAATGGCAG ACAGAGTTAG	5460
GAGTTGAGGA CGGCTGGAAT TCCCTCTTCT GGAACAACCA TGACCTCCCT CGTATTGTCT	5520
CAATCTGGGG AAATGACCAA GAATACCGCG AAAAATCTGC CAAAGCCTTT GCAATCTTAC	5580
TTCATCTCAT GAGAGGAACT CCTTATATCT ACCAAGGTGA GGAGATTGGG ATGACCAACT	5640
ATCCGTTTGA AACACTGGAT CAAGTAGAAG ATATTGAATC TCTCAACTAT GCGCGTGAGG	5700
CTCTTGAAAA AGTGTTCCG ATTGAAGAAA TCATGGACAG TATCCGTGTT ATTGGACGTG	5760
ACAATGCCCC TACCCCTATG CAATGGGACG AGAGCAAAAA CGCTGGTTTC TCAACAGGTC	5820
AACCTTGGTT GCGGTTAAT CCAAATTACG AGATGATCAA TGTCCAAGAA GCGCTGGCAA	5880
ATCCAGATTC TATTTTCTAT ACCTATCAGA AACTGGTCCA AATTCGCAAG GAGAATAGCT	5940
GGCTAGTTCG AGCTGACTTT GAATTGCTTG ATACGGCTGA TAAGGTCTTT GCTTATATAC	6000
GTAAGGATGG CGACCGTCGC TTCCTAGTTG TGGCTAACTT GTCCAATGAA GAGCAAGACT	6060
TGACAGTAGA AGGAAAAGTC AAATCTGTCT TGATTGAAAA CACTGCGGCT AAAGAAGTAC	6120
TTGAAAAACA GGTCTTGGCT CCATGGGATG CTTTCTGTGT GGAATTACTA TAAATATTTT	6180
TTGCAGAAAA ATTTAAATTT GAAATCGTAT AAAACAAGG GAGGACTGTA TAAAAGACAG	6240
AAATCCTTTG TTTTATATA CCAAAGTTTA TAACTTTCA TTCTTGAAAT TCAATTAAC	6300
TTACAAATTC CCACTATTAA GCAGAAAGAA GATGAACATA AAGAAGCGTG TCCTTAGTGC	6360
AGGCCTGACT TTTGCATCTG CTTTGCTTTT ACCCAAATCA TTCATACCTC TCTCAACTAG	6420
ATGTAACCTA CAAAACCCCT GACCTCATGA GCCACTTTCT TCCTCCTCAT GAGGTCAGTT	6480
TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT TTCCTCAAAA GGGCAGACTC	6540
CTCCCTTGGT GCGTCACACG ATTTTTCAT CTCGACTGTT CTTTAATGCA TCATTAACGA	6600

681

CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC AGGTTGACTT TTCTAATCCT	6660
AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA CTAGACAATT TGAGGAGCTG	6720
CTTGCGTCCT GTTCGAACAC ATTTTCCGG	6749

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

(A) ~~LENGTH: 1842 base pairs~~  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTACCCATG GACTTTGAGG CATTCAATTGT TCCATCTTCT AGTGGCGAAT CTTTGGATAC	60
AAACGATTCA ATTCACCTGG ATAGTGAAAC TCTCCCGCAA ACATTTTCTT GGTAACTCA	120
ATCCAGCTGA TATTCTTTTC AGCCAAAATA ATGGACAAGT TCTCCCAAAA TCGTTCAGCC	180
ATATTGCTTC TCCTTTAGTT AGATAAATAA TGTGTTTGGC CCATGTAAAT CAATTGTTTC	240
GTATCTCTTG GCAATAGAGC TCTAGCCTCT TCCAAATTCA GACTTGGATA AACTCGCTTA	300
TTTGAAACCG CAAGAGGAAG TCTGATGGTT AGTTCAGGAT TTTTAAAAAT TATCTCAACG	360
AAATCCGTGA ATCTTAGATT GTCACGGTTC TTAAATCGTA ATAAATTGGG AGATAAAAAAC	420
TCAAAACAAT CTGAAGAATA GCTCATCATC TCAATTAATT TGTCCTTTGT CATTTCAGAA	480
ACTGAATGAC AAGATACCTC TATGCCATAG TTTTGAAGA AATCTAAAAG AAGTTGATTT	540
CTTTGTCTAT TTTTACTTAG ATAGAGATCA ATCATGGGAG ACCTCCCAAA GATTCGGTTC	600
CATTTGATAT TCTGACACGA TTAAGGAATC TAATAAATTA AGGAATCTAA TAAATTTGCG	660
AAGTTAATCG GTTTCTTGTC TTCATCATAA GCTTTTACAG TTAAGTTGGT TGTAAGTATT	720
CCCTCTTTTC CCTCGGCTCG ATAGCCTTGT CCATATAAAA CAAAAACGAG ATTTTGATGA	780
TCATCTACAA AGGCATCAAC CCCATTCTTT ATGTCTTGAC TTTCAAGGAA TTCCATAACG	840
TTTTGAAGAT AGGATTCGTA AAATAGTGGG TAGTTATGTT TTTTATGGTA ATCATCTAAA	900
AATGTCACCT CAAACTCACA TGGAGAGTAA TTTTGACTTT GAACAGCCTA AAAGTGCCAT	960
CAAATTTGAA TTGGAATAAA TCAAATAAAT AGCCCCATCC TCATCAATCC AACCTTTGCT	1020
CAAAGACAAC TCCAACCGAT CTTTAAAAAC TGAGTAAACC ACCTTAACCT CCAGTTTCAT	1080
ATTCTTATAC CGTTCACCTC CAAATAAAAG TTTGGGGAGC TTATAATAAC GCTCTGATGT	1140
CTGATATTGA TTAGCGGTAA TACGCTTCAT TATTGTCCCT CCAAGACTAA AATTCCAACA	1200

682

TTTCCAAATT CATCAAATCG GATTAAACCT ACTTGTTCCA TTTCATCAAC TAAC TGAGTT	1260
GCTTTTACCC AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC	1320
TCATGAGCCA CTTTCTTCCT CCTCATGAGG TCAGTTTAC TTTCTGCTGT TCCAGTATCG	1380
TTTTTCCCTCG CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT	1440
TTTCATCTCG ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA	1500
AGGAACAGGA AGATTCAGGT TGA CTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC	1560
GGAATAGGCA TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCCTGTTT GAACACATTT	1620
TCCCACCACG TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC	1680
TCCAGCTAGA TGT TTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATA CG	1740
AACTTCGTTT TTGATTAAAG TTGAACTATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT	1800
CTCCTTGATG AAATTCTCGG CTTGACCACG TCCACGATAA AG	1842

(2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19390 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCATCTTTAT CTCCTCGAAA TTTTCTAATA TAGCCATTAT AACAGAATT TGTGAAAATT	60
CCTATTATAG TAAATCACTA TTTCAGTATA AAAAGAAAAA ACGAATCAGA CGATTGCTC	120
TTCTTAAAAT CTGAAATAG CTTTCCAGAA AGGATTAGCC GATTTTTTGC AGATTGAGCA	180
CTGCATCGTG ACTCATCAAG ACTTGACCAT ACTCTGTAA GACTGAGCGA CTGATATCAC	240
TATCGTCTGC AAACGCGGC ATACGGGCCA ACAGCCAAGC TGGATATGGG CTTGGATGAT	300
TTTCAATATC CACTAAAATG GTCAAATAAT AGCGCTCGTT CATTTTGTAG AGTTCAGAAG	360
TTTCCATTTC AAAAGTCACT GTCTTGCAA AAGCTACCAA GTCAGCCAAC TTAGCAAAAG	420
AAAGGATGTA GTAGATGTAA GGTTCTTCT TACTCTCAGC TTCTTGTTCA GCCTGCTCTT	480
GCTCTTCTTC CTTGACTTCA ACTTGCTCAA GAGATTGAAT GGCTTCGATA TCATCCTTGG	540
TTTTGTCTGC GATGCTTTTT TCCAGGGTTT TGATAAATTC ATCTGGAGAC ATTTGAGCCA	600
ATTCTTCCAT ATCTGGCAA TCCGATAAGT CTCAAAATC TAGATTTTGG TCAATCTTTG	660
ACTTGGTCAC AAAGACATCT ACCTTATCAG GTTTTGGAGT CACACGGAAG CTCAACATGC	720
CTGTATCCAG AAAGCTATCA GGCATCTCTA GCTCATCCAA GATAGCATAA AAGA ACTCTT	780



683

CTGTTTTTTC TTGAGGAACG AGAAAGTCAG CAATCTCCAT TCCACGATCC ATCAAATCCT	840
CTAAAGATAT CGTGATTTTT AAAGTTGTAT CACTAATTTG TTTCATTTTC ATTGCTAGTA	900
ACCTCATACT TTCAGTTCTA TCTATTATAC TAGATTTTTA CGATTTTATC AAAAGAAGGC	960
TCCTCTATAC GGATAGATTT TCCCTAGGGT CTTTCTATAG GAGACTCCAA AAGAAAATTT	1020
CTGCAGACAG ATAGAAAAAG CCTTCAAAAT CGGCTAAGAG CCGACTTTGA AGACCTTATA	1080
CATCAGAATA CTTATAATTT AAAGGTTGCT ACACCGAGGA TAGAACGATT TAAGTTTCTG	1140
AGAATTTGAA GACTTTGCTC AAAATTTCTTA TAACGAGTCA CTCCGTA CTC TCAACAAGA	1200
AGGACTGTAT CTCTTTCCAA AAGAGATGAT ACATCCTGTA AATCTACAAA ATGCATTCTT	1260
TTTAAAGCTT CTGACTCTG TTTCAATTTA TCTAAGATAG CTTTATTTGA GCTAACGATC	1320
GTCAATTCTT GTCCAGTATT TTTGTATGAC AAAACATCTG CTAGGTTAGC AATTGTTGTA	1380
ATCTCTGTTA CAAAATCAAT TTGATACTGA GAAAAATCAC CTACTCTATT GATTGTTGGA	1440
TTAAAGAGAT AAATAACAC ATTTCCCATC ACAACCAAAA TCACACAAAC CACTCCAATA	1500
ACAACTAAAC GAAGAATCAG ATTTTTCACA TTTAAGCCAA GCGCTGTTTC ACCATTGCG	1560
TTCAATTCTT TAGAGTTGAT GGTTCACAGT TTTTCAATTT TCACATTTGC ATAGGCATGT	1620
TTAAATTTCT CAATCAACCC ATCAATTTTT TTCTCTAACA AGTTATTGGC ATCTTTACTT	1680
GATGTCAAAA TTTTCACACC AACCCCTGCA TCGTCAATCA TATAGTAGAC GGTCAATTTT	1740
TTCCACCAAT AGTCATTTCGT TGAATTTTTT AAGGTTGTTT CTGTCGTGTC TAATTCACTG	1800
GCAATTTTTT TCAACTCACT GGGTTCTACA TCATTGAAAA GATAAGCTCC ATTCAAATTA	1860
CCATCAATCA ATTTCCCATC AAAATCACTA TAACCACCAA TTTGATGATT CAAAATCGTT	1920
TTGTCCGACT CTTTGGGAGG AGTGATTTTA TAGATAAGAT AAGTTGAATA ACTTGTTGTA	1980
TCTTTGACAG TGTTTTTATT CCTAACTGCT TTAATTGTAA ATGGTACAGC AATGAGAGCA	2040
AATAAAGCGA TGAGAGCTAA AATATTTGCT TTTCGCTTTT TATAAAGATT TGCAAACAAA	2100
TCAGCTACTG AATAATGTTT AAACATGATT TTTTCTCCT TTGTTTAGTA GATACTAGTT	2160
TTCTTTTGTA AGCATTTTTG CTACAAATAT AATCACAAGA ACAATCCCC AGAATTGCAT	2220
TGTAATAAAA TTGAAGAAAC TTTCTGAAAA GCTGCTTCTT GGCATAAAGA ATAGATTATT	2280
CAAGATGAGT AGGGATAAAG CAAATAGGAT TGTCCTTGAG CGATAGGCTA CTTGCAGCAT	2340
GGCTATAAAT AATACGCCGA GTAAGAACT AAGCAGAAAG ACTCCAATCA TACCATAGTC	2400
GGTATACAAC TCCATGATAT AACTACTTCC GATACCATGC CCTTTCAAGT ATTCCTTGTT	2460
CAAGACAAGA TAGGATAGAT TGTGGGCATA ACTATTACTA TCAATAGCTA GTTCCACACT	2520

684

ATTGGTTGTA TGTTCAAAGG CTTTTCCTCC GAAAAATGGCT CCCAAACTCC CCCTTGCAAA	2580
ATAATCAAGA ACAGGACCAA AAGTAAAATT ACGGAAATCT CGGTAAGGGA GGCTACTGTT	2640
AAATAGAAAA CCTCGAGCCA GAACACCAAA ACTAGTCCCT TGTTTATAGA TAAAGTCAAG	2700
TAAGATATCC CAGAAACCTG TATGGGAAAC TTGGACATTA TCCCGTACAT AATTGAGTAC	2760
TCCCATCGCT AACATGAGAA TAGGAGAACC TACAAAAATC GCTAACTTTT CTTTAAACCC	2820
AATCCATTTT CCTTTTTCAG TTTGCTCCCG CATAAAGTAA TAAACAAAAG CAAATmAAAT	2880
ACTTAAAATA AAGGGATTTC GTGTCCCAAT TGCCAAATGA ATAGTATTAG CTGCAATAAA	2940
GGAGACAAGC ACTGCTGTGG CCTGCAATTT CTTTGGCTTG GTTGCCAGAT ACATACACAT	3000
TGCATAGACC GTAAAGCTAG ACAAAATSTA GGTAAAATAA GGCAGTTTAC TTTCAAAATT	3060
TGCATAGTAG GCATAGTAGG AAGTCTGCAA ACGATACAAG AGCCGTTCAT ATAACCGAAT	3120
GAAATAGAAA GGATAAGTTA GAAGAAAAAC TCCTAGTGAT ACAAAGCGTA ACCGCTTGAT	3180
ATAAACCTCT TTTAGAGAAT TTCCTATATT TGCTACTTTT ATTTTCTTCC TAGCTATGAA	3240
GTAACGAGCC AGAATGCCTC CTGTGGTCAA GCCCAGAATC GAAATCATGA CAACTATAAA	3300
GGCAAAACGA TAGGCTATTG GATGATAGGT ATCCAAAGCA CCATCCCTAA AATAATCAAT	3360
GGTCGGTCTT GATACCAGAA ATACAAAAAT GGTAAATAG AAAATAAAAT GGATTAAGTA	3420
ATACTTGATA TCATTCCAAC AAGCAATTAA GCTACTAACC AACAAGAACA ATAAAGTAGA	3480
AAGTAAGCTA ACATTATTAT TATTAACAG ATACACAATT CCACTTACTA GCGTCAAGGC	3540
ATAACTGACT ATGGTCAAAC TAAATAATAA TCGTTTCCCA TCAATCACTT GGTCACCCCC	3600
GTTCTAATGT AATTTTITAG ATTTTTCAT ATTTTTCAGT AATAAGAATC GATATAAGGA	3660
AATATTTATG AATAGGGCCA AAGCACTAAT TCTTCTCCCC TTACGGAAAA TTGGATTCTT	3720
AGAAATAGCA AAGGCATGGC CTTTTAAAAA ACGATGAATC TGAGAATAGG CTTCAAACCTG	3780
TTTATACTGA TCATCTAGCA ACATCTTATC CAGAATAAAG AAGTGGGCAT AGGCCAATCT	3840
GAAAAAAGCG ACCTCTTTCA AGTCAGGATA GTTTTTCACA ACTTCATTAT AAAACTTTTG	3900
GTAGATATCA ATATAGGCTA AATCCTTCTC TGCATAGGGT TTGGTCGTAA TACTATCCCC	3960
TCTATGGAAA TAGTAATAAT AGGGTTTAGT ATTAACCACA TACTTCTTGG CCAACTTGAT	4020
TAAATCAAAA TGGTAATAGG CATCTTCGTA AATCAACCCC TTAGGAAAGG ATAGGGCAGT	4080
TGCAATCTGT CTCTTGATTA GCTTATTGCA AATCGTCCCA GGTATTTTTT CACCTATGAG	4140
GTATTCCTTT AGAAATGTTT GAGAATCACA GACAAAATAG TCATCCTGAT TGGCTGACTG	4200
TGGGCTTTCA TCATTAGCAT AGACATTCAT GACACCACAG CTCGAAACAT CCGCATCTTC	4260
TTGAACATAAT TGCTCATATA AGCTCTGAAT CATTTCTGGA TGGATATAAT CATCTGAGTC	4320

685

AATAAAATC AGATAATCCC CGTGAGCCTG CTTTCATCCCA TCATTTCCGTG CTTGCGACAA	4380
TCCTTCGTTC TTTTATGAA GCACTGACAC CCTGTCATCT TGTTCAGCGA TTGAATCACA	4440
CAAGCGACCA CTTTCATCTG TTGCACCATC ATCAACAAGA ATAATTTCCA GATTTTGATA	4500
GGTCTGCTTC TGAATGGAAG CTATCGATTT TTCTAGGTAC TGCGCCACAT TATAGACTGG	4560
CACAATCACA CTAATTAATG CAGTTTCCAT GCTACTCCTC TAATAGTTTT TCTACTTGT	4620
CGATTTGTTT TGTAAATGTA AATTGTTGAA TGAATTGGCT AGCCTCATCG ACATCAAAGT	4680
TTGAGGCAGA AGTCATGTAA TTAGTAATCG CCTGAGCTGC CTCTTGATTG CTCTCAATGA	4740
TTTGTCCAAA TCGTCCTTCT TGGGATAATT CCTCAGCCCC TCCAACGTCC GTAGAGATAA	4800
AAGGGAGTCC CAGACTCAAG GCCTCCACAT ACACTCCAGG AAAACCTTCT TGTTTAGACA	4860
TAGACAAAAG AACTTTCGTC TGAGATAGAT ACTGATAAGG ATTTTTTTGA TAACCAAGGA	4920
AATGTACATA GTCCTCAATC CCATACTCTT TGACTCGTTT TTTTCAGTTCC TCTTCCATAT	4980
CACCAGCCCC GATAAAATAG AGATGATAGT TTTTCCCTC TTGGTGTAAT AATCGTATCA	5040
CTTCCACTAC ACGGTCAGAA CCCTTATTTT CCTCAATCCG TCCGATAGTA CAGATACTTT	5100
GAGGAGCAAT CTCGATATCG ATCTTCTCTT GAGATTTTTC TAGAATAGTC TGAAAATCAT	5160
ATCCATTGTA GATTGCTCTGT AATTTAGAAG TATAATCTGG ATAAACTTCC TTGATAGAAT	5220
TGCTGGTCTT TTTTGAAATC CCTACAATTG TATTCGCAGC ATCCAACCTG CTTCTATGTG	5280
ATTCTCTTTT AGAGCTATCC TTAAGAAGTT CTTCAATACT TCCATGAATC CAAGATATCT	5340
TCTTGACTTC TCTTCTTTTA GAGAACAACA GTGGTGGATT CATAATGGTA AAAGAACTT	5400
CAACATCATA ATCATCTTTT ACAAGCAAAC GACGAGTCAG TCTTGGAATA TAAATTTCTCA	5460
TTCTCCACAA AAAAGCTCGT AACCATCTGG TTTGGCGATA ATCTTGAAGG GATTTTAAAA	5520
TGCGTACATG CTTTGAACA GATTCATATC CCTTGTCAAA GTGCTCCATT TCAAGAATAT	5580
CAATATCATA CTTTCTGGA TCCAGATTTG AAACAATGGT TGATAGAATC TTCTCTGCAC	5640
CACCTCCAAG AGAAAAAGAC CACATAAAAA ATAAGATTTT TTTCTTAGCC ACCATATTCT	5700
CCCTTGATTT CTGTATAAGA CTTATCCATA TCAGCGATGA CAGCATCATG ATGCGGTACC	5760
TGCTTGCTCG CTGGTGGAGG CGTCATATAA TCCCCAAAAG CAGTTCTGAG ATAGACATCA	5820
TAGCCGATTG GAATAGGCAT CTCTGTTTCT TCAAATGGCA AGAAAAGATT GTCTTCAAAA	5880
GATGTGATTG GGTACTTGTT TCTCATGTAG CCAGGACCTG AGCATAATTC TGTAATGCCA	5940
TCACAATCAG CCAAATCATA CTTAGTCATT TCTTCTCAG CTTTTTTCCA GATGCGATAA	6000
CGGAGAGATT TTGGAGTCAA ACCCAGTAAA ATGCGACTTC CCCATTTTCAT GAGATCACCA	6060

686

TGCTTTTCTG GAATAGTTTG CGCACAAAAG AGTGAATAAA TCAAGGCCCA ACGAACCTGT	6120
TTTTTCCGCT CAGCTGGATT TTTCGGATAA TAATCCAAAG GCAAAACATC CAAGGCCAGA	6180
CCATGTGGCA AATCCAAATC CTGCTGATAA GGCTTGATAC AGGTGGTTTT CTTGTCACGA	6240
ATGGTAATAA AAAGATTACG ATCAACAAAA TCCTTGAGAC TCTTTGACAA GAAATAACGT	6300
TCATCTGCAT AACGAGGCCA TAATTCTGCT AATTTCTCAT AATCTTTACG AGGCATAAAA	6360
AAGTCTAGGT CGTCGTCCCA AGGAATAAAT CCCTTGTTTC GAAGGGCACC AATAGCGCCT	6420
CCGCCACAGA GATAACAGAG CAAATCATGT TCTTTACAAA AGGCCACAAA ATATTCAGCC	6480
ATCTCCAGAC TACGAGCCTG AATTGCTTTT AAATCAGTCA TATTGTTTAT TATTCTTTCT	6540
ATCGTATCGT TTCATTATAC CACAAACAAG GGGTGAAAAT CTATTGCAGA CTGTAAAAAA	6600
TCAAAGCCTG ACTGCTATCC AAATAGCTAT CAACTTTGA TTTTCTGTC TTATACTCTT	6660
CGAAAATCTC TTCAAACCAC GTCAGCTTCA CCTGCGCTA GGTATAGGTA ACTGACTTCG	6720
TCAGTCTTAT CTACAACCTC AAAACTGTGT TTTTAGCAGC CTGCGGCTAG CTTCCTAGTT	6780
TGCACTTTGA TTTTCATTGA GTATTATCTT ATCTTAAGCC CATTGAGCG AGCTTGGTTT	6840
GATATTTGTT TTGATCAACC AGCAGGCCCA AGCCCCATA AACATCATAG GCATCTACCC	6900
AGTCACCCAG TTCTGGAATC GTCAATTTTT CAATACCATT TTTTGCTCCA TCCAAAACAG	6960
ATAAACCGTT TGTTAGGAGG AAAGTATAGG GTACGTTGGT TGAGGTCATA GCAAAACCT	7020
TTCCAAGAGC TTCAGAACCA GTGAAAAGTT TAGTGGGATC TTTAATTTGC TCTAAAATTG	7080
CTGTAAAAAC TTGTTGCTGT CTTTTGTAC GCGCGTAATC TGCCCTCATCA TCATCACGGA	7140
AACGAGCATA ATTGAGCAGG GTCGAGCCAT TCATCTGCTG TTTTCCGACT TTAATGGTTT	7200
GGGTTGGAGA CTCAGTCTCG GTAGCGTATA AATCATCTCC GACTGTAGCT TCTGTTAGGG	7260
GACGCCCATT CAATGTTGAA AATTGAGCAT CAATCGTCAC CCCATCAGGG AAAAGCGTGT	7320
CAATCGCTGT GGCAAAGGCC TGGAAATCAA CCAAGGCGTA GTACTTAATG TCCAAGTCAA	7380
AATTATCTTT CAAGACTTGG CGAACCATT TCGCCCCCTT TTGCCCCCTCT TGTTCCTCTA	7440
ACTCGTAGGC TACGTTTAAC TTGTTATCTG TCTGTTTTCT ACCATTAATC ACTTGACTAT	7500
AACCATCTAT ATAGACCAAA TTATCAGCA TGAACTGAC TAGCTTCATT TTCTTATCTG	7560
AGCCCCGAC ATTTAATACC ATAATAGAGT CAGTTCGTGT CTCAACACTG TTCTGGCCGA	7620
TTCGACCATC AGTACCCATG ATTAAAATAT TAACTCCATC TCTAGTGTC TGACCATTAA	7680
AGACTTCTAC TTGAGCTGCC CGGGCATCAG CAGTTTTCTT TCGGCTAGCA TCTTGTAAC	7740
CACGCAAAAA CATGAATACC ATGGCCAAAG CCACACAGAC CAAAAGTGAA AAAATCACCA	7800
TAAAAATTCG TTTAAGACGG AGCTTCCGTC TTTTCTTTT TGGAGGGAAG GAGAGTGCTT	7860

687

GTGATTTGGA TTGTGAGCGA CTCCGGTTCG CATAGCTTGG TAAGTCAACC TGCTCTTCTC	7920
TTTCTTGTTT CAAGCTAGAG CTACTATTTC CCCTAGCAAG AGTTAGCTTT TCTTGCAAAT	7980
AGGCCAACTC ATTTTTTTCT CTCTCATTGA GATAGTGAAT ATTTTTTAGC AAATAATCAT	8040
AACGCAACTG CTCATGATGA CTTAAGGGAT TTTCTTTACT CATCTTCTCT CCTTTCCATG	8100
GTCTGATATT GGATAAATAG GATAGGCACC CAGAATTTTA TACTGGATTG CAATCGCTTC	8160
TAATCTTTT TGGGCAAAGT GGACCAAGTC CTTATCGGTA TAATCCACAT CGATAATGAA	8220
AAAGTATTCA CCCAGTGCTG TCTTGAGTGG ACAACTTTCA ATTTTGTGCA AGTCAATTCC	8280
TCGCCAAGCA AAGGTCGACA GGGCCTTATA AAGTGCACCT GGAAGGTTGT CAGGTAATGT	8340
CAAGGCCAAA CTCATCTTTT CAGTTTGTGC TTGCAAGGGA ATACTAGGCT TTTCAGCTCC	8400
TAGAACCCAG AAACGTGTGA AATTGGCTTC CATTTCTCTGA ATATCCTCGG CAATCAGTTC	8460
CAATCCATAT TCTTCAGCAG AACTTCTAGG TGCAACTGCT GCAAAGGGCT GGTCTGGATG	8520
TTCGGAAATA AAACGGGCCG CATAAGCTGT ACTAGCTGTT ACCTCGATTT GAGCCTCTGG	8580
ATATTGTTCA TCGATGAATT TCTTTCCTTG AGCCAAGGCC TGTGGATGTG AAAAAATCTT	8640
TTCAATCTTA GTATGGCCTG GAACCACCAT CAACTGCTGA TGAATAGGCT GAACGATTTT	8700
TGCTACTGCT TGGATGTGAG CCTGATGAAA AAGATAGTCC AAGGTTTCAT GAACACTACC	8760
CTCAATAGAA TTTTCAACTG GCACCACAGA ATAGTTCACT AATCCTTGCT CATAAGCCTT	8820
GATGACATCT GTAATGTTGG CAAAAGCCTG CAATTCCTCA TGAGGAAAAG CTGTCTGCAC	8880
AACGTGGTGT GAAAATGATC CCTTGGGACC TAGATAAGCA ATTTTCATCT TAGTTCCTCT	8940
ATAATTTCTT CTGGGCTTAG CTTGGTCACA TCCAAAACCC GACTAGCCAC TTCCTCATAC	9000
CAAGCCTGTC TTTCTTGGA AATAGCTACT AGTTCTTCCT TGCTATTATT TAGAAAAAGC	9060
GGTCGCTGAT TGTCTTATC AGCTGCGATA CGTTGGTAGA GGGTTTCAAA ATCTGCTCTC	9120
AGGTAGATGT TATCTGTATT AGTCTTGAGT AAGTCACGAT TTCTCTGAGA AATAACCACT	9180
CCTCCTCCAG TTGACACGAC TTGGTCTGTT TGTAGTAAAT CAGCTAGGAC TTCTGATTCT	9240
ACCTGACGAA AGGCTGTTTC TCCCTTTTCA GCGAAAAAAT TCGCAATGGA CATACCTAGG	9300
CGATTCTCAA TCAGAGCATC CATATCAAGG TAATTAGGGT CCAAGCCTCT TGCAATAGTC	9360
GATTTTCCAG CCCCCATAAA CCCTAATAAC ACCTTAGCCA TGAATCAAGC TCTCCAAATC	9420
ATCAAAGAAA CTAGGATAGC TGGTATTGAT GGCTTCTGCA CGGTCAAGCT CCACCTCTCC	9480
ATCTGCAACC AAGAGGGCTG CGATAGCTGT CATCATGCCG ATACGGTGGT CACCAAACGT	9540
ATTGACTCTA GCACCGTGAA GAGCTGATTT TCCTTTGATA ATCATCCCAT CTGCCGTAGG	9600

688

AGTAATATCT GCTCCCATAC TATTTAAGGC GTCTGCCACA ACCTGAATAC GGTCTGTTC	9660
CTTGACCTTG AGCTCCTCAG CATCCTTGAT AACTGTTACA CCTTGGGCTT GGGTCGCAAG	9720
CAGGGCAATA ATGGGCAATT CATCAATCAA TCGTGGAATC AAAGCGCCAC CAATCTCTGT	9780
TCCTTTCAAG TCAGAAGACT CAACAATCAA GGTAGCAGAT TTAGCGACTG GATCGATTTC	9840
AGTTATTTCC AATTTTCCAC CCATGGCAGC AATGACATCA ATAATACCGG TCGGAGTTTC	9900
GTGTATCCCC ACATTCTGCA GCACTAGACG AGAATTGGA GCAATCAAAC CTGCGACTAA	9960
CCAAAAGGCT GCACTGGAAA TATCTCCTGG TACGACCACC TTCTGTCCTG TCAATTTTGT	10020
TGGCCCTGG ACTGTGATTT TCTTACCATC CACACTTAAA TGACCACCAA ATTGTTTCAA	10080
CATATCTTCA GTATGATTAC GGGTGTACTC TTTTTCGATA ATAAGTACT CCCCCTTAGC	10140
TTGTAAGGCT GCAAACATCA AGGCTGACTT GACTTGGGCA GAGGCAATTG GCAACTCATA	10200
ATGAATAGGT CTTAGGTTTT TCGTCCCTTT TAAGCGAAGG GGAGGCAAGT CTCGTTCACT	10260
TTGCCCTGAA ATGCTGACGC CCATTTTTTT CAGTGGGAAGG GTCACACGGT CCATAGGACG	10320
TTTGAAAGA CTATCATCTC CAAACATCTC TACTTCGAAA TCTGCACCAG CAAGGACACC	10380
TGAAATCAGG CGAATCGAGG TGCCAGAATT TCCCATTATTA AGGGCATTTC GTGGCGCTTT	10440
TAAGCCAGCC ATGCCTACAC CTTGAATGGT AATAACCCCA TCTTTATCCT CAATTTCAAC	10500
ACCAAGGTCA CGAAAACTT GCATGGTCTG AAGAAGCTCT TCACCTCGCA GAATATCATA	10560
AACCTTGGTC TCACCCTCAG CCAAACCTCC AAAGATAATG GAACGGTGGC TGATAGACTT	10620
GTCACCTGGG ACGCGGATAC TACCATGTAA ATGGCGAATG TTTGTTTTTA GTTTCATACT	10680
GGACCTCATA CTTGCAATAC TTTTACCTAT TTTATCATAA AAAGCCAGAA ATTCCCTAAA	10740
AATTCCTGAC TTTAGGATCG TTCTTTTCTT ATTTTCAGCAA TTCTGAAACT GGTTCAAAAA	10800
CAATTTTTTC AATATCAGAA AGGTAAATGG CCAATTGTTG TTGCTTGGTA AAGAATTCTG	10860
ACAAGAGGCT ATTTCCCTGA ATCTGTTTAC CAAAGCCTTC CATCTTAGCT TGGAAGGACG	10920
CATCTGGCAT TTGACCTGTC TGTGCTAGTT TTTGAATTTC CTCTTGAAAG GCAAGATAAT	10980
CTGTAAAGAT TTTGCTTGCC TCAGCATCTG CTGCAATCGC ATCTTTAGCT GCTTTAACAG	11040
CCTTGATTC TGGTAATCCG CGTAGACCGC GACTGAGTTC GTTTCGACTA TCGTAAATAT	11100
TTGACATGTT CTTCTCCTTA TTTGATGACG ACTGTATAGT CAGTATTTTC TGTATGAGA	11160
TGCTCAGCTC TTTCCAAGTC TTGAGCATTT TTAATGAAA TTTGTAGGAT TCCGTGAATA	11220
TCCTCAGGAT TTTCTCGTT GATGTGGATA TTAACCAAGG AAGTTCCAGC TAGCAGTTCC	11280
AAAATCCGCA GGATGACATC TTCTTCATCA GGAACGTCAA CATAGAGGTC GTAAGAGCTA	11340
TCCACACCAC CAGGCTTATG GATTTCATG GTCTGGCGTT GTTCACGCGC TTGCTTAAAA	11400

AAGTTCCAAA TTTGCTCTTC ATCTCCCTTA CTAATGGCCT GACCAATCGC TTCCAAACGT	11460
TCCTTGAAAT CCTCAATTCT ATCCAGAATG ATCTCGCTAT TGGACAAGAG AATGGAGGTC	11520
CACATTCCTG GCTCGCTTTC CGCAATTTCG GTCATATCTC GAAAACCACC TGCCGCAAAG	11580
CGCCTTGCCA TCTCATGCTC TTGAGCATAG ACCGCAGTCT GCTCCATGAG ACTAGAAGCC	11640
AAAATATGAG GAAAATGGCT AATCTGAGAA GTGACACGAT CATGCTCCTT GGCATCAATC	11700
TCGATAAAAC GAGCATGAAG ACCTGAAAGC AGATCCTTCA TTTCCCTTAAG CGTGTCTGA	11760
CTTGTGAGGC TTGAAGGTGT AAAGATATAA TAGGCATTTT CAAAAAGATT GACATCTGCC	11820
GAAGCAGCCC CTGTCTTGTG ACTACCAGCC ATGGGATGGG CCCCACAAA GCGAACAGAC	11880
TTGCCAGCCA AATACTGCTC CGCCGCATCC ACAATGGTTG ACTTGGTCGA ACCAGCATCT	11940
GAAATAATAA CGCCTTCTCG CAAATCCAAA TTGGCCAACT CCTTAATGAA AGCAATAGTT	12000
TGTTTGATTG GCAAGCTGAG GATAATGACA TCTGCCAAAG GAGCAAACT AGCAAAATCA	12060
TCCGTTGCAC GGTCAATCAT ACCTTCTTTC AAGGCGATAT CTCTCGAAGC TTGACTACGA	12120
TTATAACCTA AAATTTTATA ATCTGGATGA TCGCGTTTGA TACCAAGTGC CATAGAGGCT	12180
CCAATCAACC CAAGACCTGC GATATAGATT GTTTTTGCCA TAGGAACTCC TTAATAGTTC	12240
TTTGATAGT CTCGGTGTTC GGCTACCGCT TCTTTTAGTT CCTCAAGATT ATCTGATGAG	12300
AAATTTTCGA GGATTTCTTG CGCCAGAACC GTTGCTACAA CTGCTTCCAT GACCATTCCCT	12360
GCAGCTGGAA GAGCAGTCGG ATCACTTCTC TCCACGGTTG CCTTGTAAGG TTCGTGGGTT	12420
TCGATATCCA CACTCATAAG AGGTTTATAA AGAGTAGGAA TGGGTTTCAT GACCCACGA	12480
ACAACGATGG GTTGCCCATC AGTCATACCA CCTTCAAAAC CACCTAGATT ATTGGTACGG	12540
CGAGTATAAC CGTCTTCTTT AGACCAGAGA ATTTATCCA TAACTTGGCT GCCTTTACGA	12600
TAACCAGCCT CAAAGCCAAG ACCAAATTCC ACCCTTTTAA AGGCATTGAT AGAGACAACA	12660
GCTTGAGCCA ATCTTGATC CAATTTTCTA TCCCATTGGA CATAGGAACC AAGACCAACT	12720
GGAACGCCTC CGACGACTGT CTCCACAACC CCACCGATGG TATCACCATC ACGTTTGATT	12780
TGGTCAATAT AGTCCTTGAT TTCCTGTTCT CGTCTTGGT TGACAATAGA AACTTCAGAC	12840
TGGGCAGCTC TTTGCTTAAT TTCAGCGACT GTCAGATTTT CAGGAACATC GATTTCTTGT	12900
CCACCAAAGA CCACGACATG GTTGGAATC TCCATATCCA GCTCAGCCAA GAGGCGTTTG	12960
GCTACTGCAC CAACTGCCAC CCGCATGGTG GTTTCACGAG CTGATGAACG CTCCAAAGAA	13020
TTTCGCAAAT CATCAAAACG GACTTTAATC CCCCCAACCA AATCGGCATG ACCTGGGCGA	13080
GGATGAGTAA TTTTCCGCTT GCTTTTAAGG CGGTCTTCAA TGTCTCCGC AGACATGATG	13140

690

TCCAGCCATT	TCTGGTGGTC	CTTATTGATG	ACATCCATAG	TAATAGGCGC	CCCTGTCGTC	13200
TTCCCGTGGC	GAACGCCCCG	AGTAAAGACA	ACCTGGTCAT	TCTCAATCTT	CATACGACCA	13260
CCACGACCGT	AGCCACCCTG	ACGGCGTCTA	AGGTCTCAT	TGATATCCTC	AGCTGTCAAT	13320
GGAAGTCCAG	CTGGAATTCC	CTCAATAATA	GCTGTTAGAC	GGGGGCCGTG	TGATTCTCCT	13380
GCAGTTAAAT	ATCTCATACA	CTCTCCTTAT	TTTACCAAGT	AGTCTTTCAT	CTCTTCCAGA	13440
GAAACTGGGT	GAATGGTCGC	TGAACCAAGC	TCTGGCACCA	AGACCAATTT	CAAGGTGTTA	13500
CCACGCGCTT	TCTTGTCATG	AGTAAGAGCC	TGATAAAGCT	TGCCAACTTC	CCAATTTTCA	13560
TAGTCAACAG	GCAAACCGAA	TTTCTGACAC	ATCTCTGTGA	TAGATTGGGT	AATGCCAGCT	13620
GGCATGAGGC	CTTTTTCCTC	AGCAACCTTG	GAAATCTGTA	CCATTCCCAT	GGCAACAGCC	13680
TCTCCATGCA	TGACCTTGCC	ATAACCGGCA	GTCGCTTCGA	TGGCATGGCC	AATAGTGTGG	13740
CCAAAATTGA	GGTAAAGACG	AATACCATTG	TCCAACTCAT	CTTCAACCAC	CATCTTGCGC	13800
TTCACCTGAC	AAGAATGTTT	AATCAAGGTC	TCTGCATGTT	CCAAAATACT	CTCAACAGAA	13860
CCATTCACTG	CCGTCAAGAG	AGCCACAGT	TCTGGATCCT	CAATCAAGCC	ATACTTGATA	13920
ACTTCACCCA	TCCCTTCAAT	CAACTCTCTT	TTTCCGAGGG	TTTCAAGAAC	AAGTGGATCA	13980
ATCAGAACCC	CATCTGGTTG	GGCAAAGGTC	CCCACCATAT	TTTTAGCAAA	TGGTGTATTA	14040
ACGCCTGTCT	TTCCACCGAT	AGAAGAATCA	ACCTGAGCTG	TCAAACCTAGT	CGGAATCTGA	14100
ACAAAGTGAA	TACCCCGCAT	ATAGGTAGAG	GCTACAAATC	CAGCCAGGTC	CCCAACAACG	14160
CCACCACCAA	GAGCAACGAT	TCCATCGCTA	CGAGTCAGAC	CTTGCTTGAC	TAGAAATTCA	14220
TAGACTTTCT	GAACAGTAGT	TAAATTCTTT	CTTTCTTCAC	CTTCTAAGAA	ATCAAAAACA	14280
GCTACCTGAA	AACCAGCATC	TTCTAGGCTG	AGCTTGACCT	TCTCTGCATA	GAGAGAGGCT	14340
ACATGGTTAT	CTGTCACAAT	GACTACCTTT	TGCGGTTGCC	AGAGTTCTCG	CAACCACTGA	14400
CCAGCCTGGG	CCATACAACC	TTTTTCAATC	TGAATATCAT	AAGGATGGTG	AGGAATATCG	14460
ATTCTGATTT	TCATAGGAGA	GTCTCCCTTT	CTTTATTGGT	ATTTTCTGT	TAAAGACTGC	14520
CAAATCTCTT	CTGTCGGCAT	TTCTTGCCCT	GTCCACAGTT	GAAAAGCTTC	TGCAGCTTGA	14580
TAGAGTAACA	TTCCAGACCC	ATTGACTGCT	GGATTGCCCT	GACTTCTAGC	CCATTTCAAA	14640
AACGGTGTTC	CAAAGGGTTG	GTATATGATA	TCTGCAACTA	AAAGAGTTTC	TGGTAAGACT	14700
ATGTTTTTCAG	GAACAGGAGA	GGATTGGCCA	TCCATGCCCA	CACTGGTGGC	ATTAACTAGC	14760
AAATCCGACT	CGGCAATCCT	TGCTTGACGT	TCAGAAACAT	ATTCTAAAGC	ACACAAATCC	14820
ACTTTAAAC	CTGTCTGCTC	CTGTAACCTG	TCTAGGTAAG	GTCTTGTTTT	TTCCATAGAA	14880
ACGGAACGAA	CAAAGACCGA	AATCTGACTG	ACGCCATCCA	AAATAGCCTG	TGCCAAGATT	14940



691

GATTTAGCCG CACCACCTGC ACCCAGCAGG GTCATCTTTT TACCTGAAAT TGTAAGAGAA	15000
GGCAAGCACT TAAAAAATCC CTGCCATCT GTATTATATC CAATTAAATT GCCATTCTCA	15060
TTGACAACCG TATTAACCGC ACCAATCAAG CGCGCTTCAT CGCTCAGCTT ATCCAAATAA	15120
GGAATCACCT GCTCCTTATA GGGCATGGAC AGATTGATGC CAAACATCTG GTAGCGACCA	15180
ATATTGGCCA CTGTTTCTAC CAAGTCACTC GCTTCAATCT CCCAAGCCAC ATAAGCACCG	15240
TTGGTAGCTG TCGCCTCAAA GGCTCTATTG TGGATGAAGG GAGAAATAGA ATGCTTAATA	15300
GGATTGGCAA CAACTGCAGC TAAACGTGTA TAGCCATCAA GCTTCATCCA AAATCTCCCT	15360
GATTTTTTTC ATGCTAGCTA GAGAAATCTG CCCAGGGGCA CTAACCTCAT CCAGACTGGC	15420
AAAAGACCAA CTCGAACCAG TCACATCCGC AGTGATACGA GAGACCTTGC CCACCTTACC	15480
CATAGAAATG GTCACATATT CCTGTTCAAG ATTGAGGGTT TTAAGCCTC GTGTATAGTT	15540
CATCAAGTCT AAGACATCCT GCTCCGTGTG AGCCATCACC GCAACCTTAA CAAGTTTGG	15600
ATTTAGGATC GTCAACTCTG ACAAGATTTC CATCATGTTT TCAGGTGTTT CTGGAAATT	15660
ATGGTAACTC AAAACAAGAT TTGGGAAGTC CAGCATTTCC TCAAAAACAT CCTTGTAGCT	15720
ATAGTACTCA AAATCAATAT AGTCTGGTTG ATAGAGTTGC GCAACTTCCT TGATTAGATG	15780
GATATACTCT TCTGGAGAAA GGTGATTTT TCCACCTTCG GAGCGAGTTC GTAGCGTGAA	15840
AACCAACTCA CGGCCTGCGA ATTTTTCAAA AATGGCTGGA GCTACCTGCA AAATCGCTTC	15900
TTTAGGCAGA TAGTCGGCAC GCCATTCAAT GATGTCGGCA TCCAGGTACC TCGTGGCATC	15960
CAGAGCCTGA GCCTCCTCTA AACTTCTTGG CATTACTGAA ACGATTAATT TCATTTACTA	16020
ACCTTCATAC TAATCACCTT GAGGTAATTA CTACTTTCAT CTTTTTTATT ATAGGCAAAA	16080
TCTGCTGGAA GACCATATTT GTTTAAAATC TGCTAACTTC TTCTTGCAA ACCTTTATCA	16140
ATTTGTTCTG TAAATTTCTG ACGGGAACA TTGGCAGCAT TGGTACTGGC AATGATAATC	16200
CCTCCCGGAT TTAATCTC AAGACTCTGG GAAATCAACT TGTGATAATC CTGGCCACA	16260
GAGAAAGTTT GTTTTTTATT CCGAGCAAAG CTAGGCGGAT CTAGGACAAT CACATCGTAG	16320
GTCAAGTCTT TGGCTTTGGC ATATTGAAA TACTCAAAGA CATCCATGAC TATAAACGA	16380
TGCTCGTCTG TGCTGAGCCC ATTTGCTGA AAATGCGCTT GAGACAATTC TCGTGAACGT	16440
TTGGCTAGAT CAACAGAAGT TGTATGGCTA GCTCCTCCCA TGGCCGAGC TACTGAAAAA	16500
GCCGCTGTGT AGGAAAACAT ATTGAGTAAG GATTTACCCA TAGCCAAGCC GTCAACTAAA	16560
CTACCGCGAA CCTCATGCTG GTCTAGGAAA ATTCCTGTCA TCAAGCCATC ATTCATAAAG	16620
ACTTGATACA GGACACCATT TTCTAAAACA TTGAAAAAGT CAGGTGCTTC TTGACCATAA	16680

692

ACATGGGCAG	ATTCATAGTC	CAAACCCCTTA	AAGCGGATTT	TCTCATAAGC	TCCTAAAACC	16740
TCAGGGAAAA	CCTGTCTAAA	GGCTTCTGAT	ATAGTCTGAC	GAATCTGATA	AACATAAGAG	16800
TTATACCAAG	AAAAGACGGC	GTAGTCGCCA	TAAAGGTCCA	CTGTCAGACC	CCCAAAGCCA	16860
TCTCCCTCTT	GATTAAAGAG	ACGAAAGGCA	GTTGTCAAAT	CATCTTGATA	GTAGGCGTTT	16920
CTCTTTTCTT	TGGCTTTTCT	AAACAACGTT	TCAAAGAAAG	CTTGATTGAA	GGCCACCTTG	16980
TCTTTGCTGA	TAAACCAGCC	CAAGCCCTTG	TTTTGCTGAG	AAAGGTAGGC	AGTCCCAAGA	17040
AAGTTTCCTT	CCTGACCCTG	CACCTCTACT	TCCTGATCCT	TAAGATTGAC	ATTCTCAAGA	17100
TCACTGGCTT	CTAGTAAAC	TAGCCCCCTTA	GCAAGCTTCT	TTTCAACCCT	TTTGCTGACT	17160
CTTATTCTAT	TCATAACTAC	CATTATATCA	AACTTTTAGA	CAATTCTCAA	AAAAGAAACT	17220
ACCTTTGCTT	TTTTACTCTT	CTTTTAAAA	ATGGTATACT	AGACTTCCTG	CAAAACTAGG	17280
AAGTAAATGT	GTAAGAATCA	CAGTAAAAA	TGCTCTTCCG	TCTTGGAGGA	GCATTTCTTT	17340
TTATCAACGA	AAATCAAATA	GCAAACATG	AAACTAGCCT	CAGGTAACT	GTGAGATTAT	17400
AGGTAGAGAG	GTTGTATCAG	CAATATGTGT	CTGTCAAAT	TAGTGACAAA	GGTAGTAGAA	17460
GAAAGATAAA	GAAATAAATC	AGCTTCAGTA	GGTATCTGGA	AAATTTGATT	TTATAGAGAA	17520
GCCTTTTGTT	ACAAACTCAA	TATACTATCA	ATAAATAATA	TTATAGAAGC	AACAATAATT	17580
ATAATTTTAC	CTATCTGCAT	CATTCTATTT	CGAACTCTAA	ATATATGTTT	TATCAAAAAT	17640
ACTTGGAACA	CACACATTAT	AGGAATTAAC	GTTTTTGAAA	TTGAAAAATA	TCCAAATAAA	17700
TAAACTATAA	ACAACAAAAA	TAGAACTATG	TTATATTTCT	TATTCAAAAC	ATTCCTCCCT	17760
ATATATTTTT	GATTACCAAT	CTTAATCATT	TACAACTACA	TTCTAACAAA	CTATAAAAGC	17820
GTTTGTGCGA	TTGAATTTAT	CAAGCAAGCG	ACCAACCAGT	TCATCTTTTT	TCTATTTCTG	17880
CCAATATGCG	TGACAGGTAA	TAATGATAGC	CAAAAATAGC	AAGAGCAAGC	AAGACGATAA	17940
GAGCTCCTAC	TCCCAAGCTG	ATGGCAAGGA	TAGGGGAGAG	AGACTGAACC	AAGAATATGC	18000
TCCCAATTAC	AAGGGCCATC	AGGATTGCAC	TATAAATAAA	CAATAAAACT	ATGGCGACFA	18060
TGCCATTTGA	ACGATTCAAC	AGGTCCGTAA	TGCTACTCCA	ATTGGTTGAC	AGATTTTTAA	18120
CGTCCTTAAA	GTAATGGTGG	CAAGAAAGGA	TGACACTGGC	AATGATCCAG	ACTACAAGAA	18180
GGTAAATCAT	CGAAATGATG	GGCAAGCCTA	GATATAGAGA	AAGACCAAGC	AAAGTCAGAA	18240
CTGGTAAAAA	GGACTGGACA	GCATATATAA	TCCAAAATTT	CACTTTCACA	TAACGAGCAA	18300
AGTCAAAGGG	TAAACTCTTA	AGAAAATCAA	CATTTTCCCT	CTCCAAGGAC	AAGGCAATTG	18360
AATGCAGGCT	GGTGATATTG	TTATTGACAA	CTGCTATAAA	GAGAGCTATA	AAAAACAAGG	18420
GTAACCAGTA	TGGAGGATGA	ATGTCTGGAA	CTATCTGAGA	ATCTCGGATT	TTGGAAATCA	18480

693

GACCGATCAT CATGAGATAA GGAAGGAAAG CACTTGTAAG AAGCACTGTA ATCACGCCAG	18540
TCCCCTGTCC CAAGAGGGTG AGGTGGTAGC GTAAAACCAT GCGAAAAAAT CCCTTTTGTAG	18600
TGGTTGAAAT TCTCTCCTTG CTGCGACGTT CTTTTTTGAC CTTCTCCTCA CTATTAAGCA	18660
GGATCACGTC ATAAAAACGA GGAAGGACCT TCTTTTGGT CAGATAAAGC AGGAAGAGAG	18720
TTAGTCCTAT CCAAGCGAGC AGACCCACTA AGCCTTCTGT CGAAAAAGGC TCCACTGCTA	18780
TTTTGTAAAA GATATGAAGA GGATAAAGGA GAAATGGAAT GTCTCTAACT TTGTCAACAA	18840
TACTTCCAAA AGTCGACTGA AGAAAGAAGA TAAATATTAA AGGTATGAGA ACTCCTATCC	18900
CAATCATCAC ATTGCAAAAA ATAGACTGAT ACTTTCTGAA GACCCTAGTT TGAGCCAAGA	18960
AATGCACTGC CACTACCATC ACTAGAGCCA CAGAGACAAA TAATAAGGTC AAGGACAGTA	19020
GCATCAAAGG CAAACCCAGC CATAGAGAAG GAGCTAGCCT AATGTAGAGG ACCAGAAAAAT	19080
AAGCTAGGAT TGGTACAATT CCAGTTAGAG CTGGCAAAAG GACAGACAGT CCTTTAGCAA	19140
TTATAATCTC TGATTCTTTA AAGGCATAGG GCCTATACGA TACCAAATCC TTACTCTCAT	19200
AAAAGACATT GTAAAAGGCC GTTAAAGAAG TTGAAAAGGC AATCACTAGT AAAATAGCAA	19260
TCATCGAGCT AAAATAAATA GGTATTTCTT CAAAAGGAAA ATGAATGGCT ATATTACTAA	19320
AACAGATGAT CATCAAGAGA CTGGAAAAAA TGTAAGAACT TAAGACTCTA GCGGAAACAT	19380
TTACTTTTTT	19390

## (2) INFORMATION FOR SEQ ID NO: 87:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 18436 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCGAGCGTCG TTACAGACTT TATCAAGATT GGACGCAAGA AGAAATTCAA CATATAAAGG	60
AAAATATGGC ACAATCTCCA TGGCATACTC ATTACCATGT TGAGCCAAAA ACAGGACTTC	120
TCAACGACCC AAATGGCTTT TCTTACTTTG ATGGCAAGTG GATCCTCTTT TACCAGAATT	180
TTCTTTTGG TGCAGCCAC GGTTTAAAA CTGCGGCACA GCTAGAAAGT GATGATTTGA	240
TTCACTTTAA AGAACTGGA ATCAAAGTTT TACCAGATAC TCCATTAGAT AGCCACGGTG	300
CCTACTCTGG TTCTGCCATG CAATTTGGCG ATAACCTATT CCTATTTTAT ACAGGAAATG	360
TTCGCGATAA AAATGGATC CGTCACCCAT ACCAGATCGG TGCTTTGATG GACAAGGAGG	420

694

GTAAGATTAC	AAAGATTGAC	AAGATCTTGA	TTGACCAGCC	AGCAGACTCT	ACTGACCACT	480
TCCGCGATCC	ACAAATTTT	AACTTTCAGG	GTCAATATTA	TGCCATTGTC	GGCGGACAAG	540
ACTTGGAGAA	AAAAGGTTTC	GTTCGTCTCT	ACAAGGCTGT	CAATAACGAC	TACACAAACT	600
GGCAAGCAGT	TGGCGACCTT	GACTTTGCTA	ACGACCGTAC	TGCCTACATG	ATGGAATGTC	660
CTAATTTGGT	CTTTGTAGAG	GAACAACCTG	TCCTTCTCTA	CTGTCCACAA	GGATTGGATA	720
AGAAAGTTCT	AGACTACGAT	AATATCTTTC	CAAATATGTA	TAAGATCGGG	GCTTCCTTTG	780
ACCCATAAAA	TGCCAAAATG	GTAGATGTGT	CTCAACTTCA	AAACATGGAT	TACGGTTTCG	840
AAGCCTATGC	AACTCAAGCC	TTCAACGCTC	CTGATGGGCG	TGCTCTAGCA	GTTAGCTGGC	900
TTGGTTTGCC	AGATGTTTCT	TACCCATCTG	ACCGTTTTGA	CCACCAAGGA	ACCTTCTCTT	960
TGGTCAAGGA	ACTCACTATC	AAAGACGACA	AGCTCTACCA	GTATCCAGTC	GCTGCTATTA	1020
AGGACCTTCG	TGCTTCTGAA	GAAGCCTTCT	CAAACCGTTC	CCAAACCAAG	AACACTTACG	1080
AACTTGAAC	CAACTTGGAA	GCTAATAGCC	AGAGCGAGAT	TGTCTTACTT	GCTGATAAAG	1140
AAGGTAAGGG	ACTTCAATC	AACTTTGACC	TTGTAAACGG	TCAAGTAACA	GTGGATCGTA	1200
GCCAGGCTGG	AGAACAGTAT	GCCCAAGAAT	TTGGGACAAC	TCGTTCTTGC	CCTATCGAGA	1260
ATCAGGCTAC	TACTGCTACA	ATCTTCATCG	ATAACTCTGT	CTTTGAAATT	TTCATCAATA	1320
AAGGAGAAAA	AGTATTTTCT	GGTCGTGTCT	TCCACATGC	GGACCAAAAT	GGTATCCTGA	1380
TTAAATCTGG	AAACCCAACT	GGAACCTACT	ATGAATTAGA	TTATGGTCGC	AAAACCTAAT	1440
GATGTCGCCA	AACTTGCAGG	CGTCAGTCTT	ACTACCGTTT	CTCGGGTTAT	CAATAAAAAA	1500
GGGTATCTAT	CTGAGAAAAC	CATCCAAAAA	GTCAATGAAG	CCATGCGAGA	ATTGGGCTAT	1560
AAACCCAACA	ACCTGGCTCG	TAGTCTGCAA	GGAAAATCAG	CTAAGTTAAT	CGGCTTGATT	1620
TTCCCAATA	TTTCCAATGT	TTTCTATGCA	GAATTGATTG	ATAAATTGGA	ACACCAACTC	1680
TTCAAAAATG	GTTACAAGAC	CATCATCTGC	AACAGTGAAC	ATGATTCTGA	GAAGGAACGC	1740
GAATACATCG	AAATGTTGGA	AGCCAATCAG	GTGGACGGCA	TCATTTCTGG	TAGTCACAAC	1800
CTAGGAATCG	AAGACTACAA	TCGTGTGACA	GCGCCGATTA	TTTCCTTTGA	CCGAAACCTA	1860
TCGCCAGACA	TCCCTGTCTG	CTCCTCTGAC	AACTATGCTG	GTGGGGTTCT	TGCTGCCCAA	1920
ACCTTGGTCA	AGACAGGTGC	CCAGTCTATC	ATCATGATTA	CAGGGAATGA	CAATTCTAAT	1980
TCGCCAACCG	GACTGCGCCA	CGCTGGTTTT	GCATCCGTAC	TCCCAAAAGC	TCCTATTATC	2040
AATGTTTCCA	GTGACTTTTC	TCCCGTCAGA	AAAGAAATGG	AAATCAAGAA	TATCTTGACC	2100
CGGGAAAAAC	CAGATGCCAT	TTTTGCTTCG	GATGATTGTA	CAGCTATTCT	GGTCATTAAA	2160
ATCGCTCAAG	AATTGGGCAT	TTCTGTCCCA	AAAGAGCTCA	AGGTCATCGG	CTATGATGGG	2220

695

ACCTACTTTA TCGAAAATTA CTACCCTCAA TTGGCTACTA TCAAGCAACC TTTGGAAGAG	2280
ATTGCTTGTC TCACTATTGA TCTTCTCTTG CAAAAGATTG AAGGCAAGGA AGTCGCCACA	2340
ACTGGTTACT TCTTACCAGT TACGCTATTA CCAGGAAAAA GTATTTAAAC ACAAGAAAAAC	2400
TCAGACCGAT TCGTCTGAGT TTTTATGATC TTAAATTTTC GAGATAGCGC TGGGCTGTCT	2460
CTAGGTTAAA GGTTTTATCT GAGATGAGGC GCTCTACTAG GGGAGCAACT TCAGATTAC	2520
TAGCCCCAGC TAGGAGAGCT AGGGATTGG CCTGTAGTTT CATGTGGCCT TGCTGGATGC	2580
CCGTACTTAC CAAGGCTTTG AGGGCTGCAA AATTTTGAGC AAGACCGATG GACACGATAA	2640
TCTGGGCTAA TTCTCTGGCA GAAGGATTTC CTAGTAGATC ATGACTGAGA ACTACACGTG	2700
GGTTGAGGCC GATAGAGCCA CCCTTAGTCG CTACAGGCAT GGGCAGGGTC ATCTCACCGA	2760
CCAATTCTTC TCTTTCAAGG TCCAGCGTCC AGCAGCTAAG ACCTTGATAG CGTCCATCTC	2820
GACTGGCAAA GGCATGGGCC CCAGCTTCGA TGGCACGCCA GTCATTACCA GTGGCAATCA	2880
AAATCGCATC AATACCATTA AAAATTCCTT TATTATGAGT AGCAGCTCGG TAAGGATCAG	2940
CCTGCGCAAA CTGACTAGCC AACGCAATTT TCTCCGCAAT CTCTCGTCTT TGATCCTTTT	3000
GGCGGCTCAA GTAGCGAAAG GCGATGCGAC AGCTTGCACT CACCAGAGAA TCGGTCGCGT	3060
AGTTGGACAG GATTCCCATG AGACTCTGTC CCTGACTGAG TTCTTCTAAG ACTGTTTCA	3120
AGGCTTCCAG CATGGTGTG AGCATATTGG CACCCATGGC TTCCTGGGTA TCGACATGAA	3180
TATAACAAC GAGAAAGTCT GGTTCCGCTT TTATCTGCTC GACATGCAGA TCACGCGCCC	3240
CACCTCCACG TTAAACGATA GAAGGATAGG CTTGATTGGC AAGCTCCAAG AGCTCCGCTT	3300
TCTTGCTGGC AATCTTCTCT TGCGCTAGTT TAGGATTAGC AACTTGATAA AGGGCTACCT	3360
GCCCAATCAT CTGTCGCTGA TGGACTTGTG CAGTAAAACC ACCTGCACGC TTGATGATTT	3420
TGCTGGCATA GCTGGCCGCC GCAACCACAG AGGGTTCTTC TGTACATAG GGAACGGTGT	3480
ATTCCTGACC GTTGACAAGT ACCTCCGGAA CCAGTGAATA AGGCAGAGAA AAAGTTCCCA	3540
CTACATTCTC ACTCAGCTGG TCTGCCACAG TCACGCTCAT CTGTTTCATCC TTCTCCAGAC	3600
TAGCTTGTCT CTCAGGACTA AGGAGCGCCT GAGCTTTTAA CAGCTCGAGG CGCTCTTGGT	3660
ATGATTTTTT AGAAAATCCA TTCCAACCTA TCTTCATTAT TTTTCAACCT TGCTATAACG	3720
GCGTTGGTGG TCGAGAATTT CAACCAAGGC AAAATCTTGA TTTTCATAGC CAGCAAACCTG	3780
GGCAGAGTTA GTTTCATCCA AGTTTACTTC CTCAAAAAAG ACCTTTTCAT AGTCTGCAAC	3840
GGATAGGGCA GTTCGTTGGT TGAGCTTGTT CAAACGGTCT TTATCCAAAT AAGCTTCATA	3900
TCCTTCAACC AATTCACCAC TGAAGAAGT AGCCACAGCT CCACCTCCGT AACTATAAAG	3960

696

GGCGATTTTA	TCCCCAGCTT	TCAAGCTATC	TGTATTTTCC	AAGAGAGACA	AAAGTCCAAG	4020
GAAAAGTGAA	CCTGTGTAGA	TATTCCTCCAC	CTTTTGACTG	TAGAGAATAG	ACTGGTCAAA	4080
ATGCTTTTGT	AAGAGGTCTT	TTTTCTCTTG	AGGCAGGCTC	TTATCCATGA	TTTTTTTCAA	4140
GCCTTTTAGC	GCTAATTTAG	GATAAGGCAA	GTGGAAACAA	ACAGCCGCAA	AATCATCCAA	4200
AGTAAGCTGG	TAGCGTTTTT	GATATTCAAG	CCAAGTCGTT	TTCAAATAT	CCAAGTATTG	4260
TTGGGTAGAA	TAGACACCAT	TTACATAAGG	AGTTGTCGAG	TAATTTGGTC	GCCAGAAATC	4320
CATGATGTCA	CGGGTCTGAG	CTACATTGTC	ATTATTAAG	GCCATCATGC	GTGGATTTTG	4380
TGTAATCAAC	ATAGCTACAC	TTCCAGCACC	TTGAGTTGGT	TCTCCTGGAG	TTTCAATACC	4440
GTATTTGGCA	ATATCACTGG	CAATGACCAA	GACCTTGGAC	TCCGGAGAAT	TTTCCACATG	4500
CAATTTGGCA	TAATGGAGGG	CAGCAGTCGC	TCCGTAGCAG	GCTTCTTTAA	TCTCGAAACT	4560
ACGAGCAAAG	GGCTGGATGC	CCAGCAAGCC	ATGCACAAAG	ACGGCCGCAG	CCTTACTCTG	4620
GTCAATTCCT	GACTCGCTCG	CCACAATGAC	CATGTCAACT	TCTTGTCTTT	CTTGCTCAGT	4680
TAAAATAGAG	TCACTAGCAC	TGGCCGCCAA	GGTCACGATA	TCCTCAGTTA	GGGGCGCAAT	4740
ACTCAATTCC	TTGAGTAAGA	GTCCTTTACT	TAATTTTTC	GGGTCAATTC	CCCTCGCTTC	4800
TGCTAAGTCT	TGTAATTTCA	AGACATATTG	ACTGGTCGCA	AAACCAATCT	TATCAATACC	4860
GATTGTGATA	TTTACCTCTG	TTTTATCATT	CATGTAAAAA	ATCGTTCTAT	ACTATTTTAT	4920
CACAAATGGC	AGTAAAAGAG	AGAAAAAGA	CTTGATTAC	CAATCAAGC	CTCTTATTGG	4980
TCATCATTTT	AAAGAATGAT	TAGTTGCTAG	AGAGTTCACC	GATATAAGTA	GCTTTATAAG	5040
CTCCATTAC	AGTTATCAGC	TCCTGGAGGA	TCAAATTTCC	TGAGTAAGTC	CTTCCCATCT	5100
CATCTACAAA	TTTTTGATAA	AACTGACTGG	TCGGAATTTT	TCTGACATCC	TTATCAAATG	5160
TCTTATCAAG	TGTTTTACTA	ACCTTCTCAG	CAATCAATTG	ATGCTCTTGC	CATCCACTTT	5220
GAAACTCTGA	GCCCCGAATA	GAAACCATGA	CTGGGATAAA	CAACAAGGTC	AGTAGATTTA	5280
CAGACAATAA	GGAAAGTAGT	AGACTTCCTG	CAAACTAGA	ATCCTAGTTC	ATGATTGATA	5340
ATACCAGCAA	TCAAATTCAT	TCGTAATCCG	AAGCGTTTAC	GATGATTTTC	ATAGGTTGTT	5400
GAAAACATTT	TAAACGTTTT	TACTTTGGCA	AAGATGTTCT	CAACCTTGCT	TCTCTCCTTA	5460
GATAGCGCAT	GTTTACAGGC	TTTATCTTCA	GCTGTTAGCG	GCTTGAGTTT	GCTGGATTTA	5520
CGTGGAGTTT	GTGCTTGAGG	ATATATCTTC	ATGAGCCCTT	GATAATCACT	GTCAGCCAAG	5580
ATTTTACCAG	CTTGTCCGAT	ATTTCTGCAA	CTCATTTTGA	ACAACCTCAT	ATCATGACTA	5640
TAGTTCACAG	CGATATCCAA	AGAAACAATT	CTCCCTTGAC	TTGTGACAAT	CGCTTGAGCC	5700
TTCATAGCGT	GAAATTTCTT	TTTACCAGAA	TCATTGCTA	ATTGTTTTTT	AGGGCGATTG	5760

697

ATTTTACTT CCGTCACATC AATCATTATC GTGTCCTCAA AGCTGAGAGG AGTTCCTTGAA	5820
ATCGTAACAC CACTTTGAAC AAGAGTTACT TCAACCCATT GGCTCCGACG GATTAAGTTG	5880
CTTTCGTGGA TACCAAAATC AGCCGCAATT TCTTCATAAG TGCGGTATTC TCGCACATAT	5940
AGAAAGCGTT ATCAATTTAT TTATCTCATT TTTCAGAAAA TTCTTTTATT TCTGTAAAGT	6000
CTACGATACT CGATGTGTTT TTATATAATG ATAGAGTCTG AGAATCACTG TTCGCTAGC	6060
CATTCCAATA GAGATTACCA AAGCCAACAT GACAACCAAG GTCGCACTTG CCAGTGCTTT	6120
ATTATAGTCC CCTGTCACAA AAAAGGCAGT TGTTCGGTAG GAGAGATAAC CTGGAACCAG	6180
CGGTGCCAAA ATGGCCAAGA TAAAGACCAC AGCAGGTGTC TTATAAAGAA TACTTAAAT	6240
CTGGCTGACA CAAGAACCA TAATGGCTGC AATGAAGGTA GCTACAATGA CATTTGGTCGG	6300
TTCTTGAGC AAGAGATAGA TTAGCCAGAC AGTCATGCCC AAAATCCCTC CAGGTAAGAG	6360
CATAGACCGT TGCACATTGA GTACGATTAA AAAAGTGATA ATGGCAAGAA AACTTGCTAC	6420
TGCTTGTAAT AAAAAGGTTG TTAGTGTCAT ATTAGTTCAT CAATACCAAG GCGACAGAAG	6480
TTCTGCCCC TAAAGCGAGG GTAATGAGCA GGGATTCAAA CATCTTACTC ATACCAGAGT	6540
TTATGTGGTT GGTGATAATA TCACGGACCG CATTTGGTCAA GGCAATACCT GGTACAAACG	6600
GCATGACCGC ACCAGCTATA ATCAAATCTG CCGTTGAAGG AAAACCTGTG TAGCGAGCCC	6660
AAAACTGGGC AATTATCCCA AAGACAAAGG CTCCAGCAAA GGCTGTCACA AAGGGAATTC	6720
GGATAAATTT TTCCACATAG AGGGAAAAGG CAAAACCAAA TAAGGTCGCC ACTCCTGCCC	6780
CAAGTGCGTC GTAGATATTT CCGCTAAACA TAACTGAAAA GAAAGGAGCA CTAAAGGTCG	6840
CAGCCAGAGT TACCTGCAAC TTAGTATAGG GAAGGGGTTT GGCTTGCAAG GCCGTCAATT	6900
GCTTAAAGGC TGTCTCTAAG TCAATCTGCC CCCCAACTAG CTGACGAGAA ATCTGGTTCA	6960
CATCGCAGAC TTTTTCGATG TTATAAGAAG AGGAGGTCAC GCGCTTCATG CGCAAAATATT	7020
GGTATTTTCA ATAGAGAAAA AGATAGCGGC AGGCATGGCA AGGACATTGC AATCCACAAT	7080
CCCTGCGAA TGCGCGATTG GAATCATGGT ATCTTCTACA CGATGGATT CTGAGCCACT	7140
TTTAAGGAGA ATAGTCCCCG CTAGCATAAT CACATCAATG ACGGCATTTA ATTCTTTTGA	7200
TTCTTCCATG CTTTCCTCCT TTTATCAACT CCCTCTATTC TATCACAAAT CCGGACTCAA	7260
AAAAATCTT TGCCATGAAA TCATGACAAA GATTGATTAC TCATTTTGAT TATCCATCTG	7320
CTTTTAAGGA GTAGCTGAAG TTGTTTTAGG TTTGTAGATT GAAATCTTGA CTCTAGTCTT	7380
ATTGAGGTCT ACCTTTTCAC CTGCTCTAGG ACTTTGTTCA ACAACCATGC CTTCTGCACT	7440
ACCTGCAGGC GCTGTCGTCA CTCTACAAC TTCTATATTA GCTTCCTTAA TCCCAACAAT	7500

698

TTGAATCAAA TTGTTCTTAG TAAACTCCAA GCTAGAACCA ATGTAACCTCG GCATGGCAAC	7560
ACTTGTAAC TTTTGTAGCTA CTGTCAAGAC AATTTGAGTA GGTTTACTCA CATCATAAGT	7620
CGTTCCGGCA CCTGGACTTT GTTTCATAAT CGTTCCTGGT TCGCTTTCGC TGGACTCTTC	7680
TTCTCTATC TTAATCAAAT TCTCAGGAAC CTCTCTCTGC TTGAGTTCTG AGATTACTTC	7740
TGTAGAGTTC CGTCCAATAT AGTCCCTAA TTGAATCGTC GTAGCTTTTT TAGCTACTGT	7800
CAAAACAATT TGAGTTGCCT TGCTCAAGTC ATAGGTCGTA CCTTCTGGTA GACTTTGCTT	7860
CAGGACCGTT CCAGCCTCAC TCTCATTCGA CTCTCTTCC TCAATTTTAA TCAAATTATC	7920
TGGAACTTTT TTCTCTTTTA ATTCCGCAAT GACATCAGAG GATTTCCGAC CGACATAATT	7980
ACTAATTTGG AAAGATTGCT TGCTGTATGA GACAACCAAA TTGATTTTCG TTCCTTCTTT	8040
TCGACCAGTT CCAGCGCCAG GATCTGTACG GATAATCCGC CCTTCTTCCA CCTTTTCACT	8100
AGCCTCTGTC TTCTCTCAC CAATCTCAAA ATTGGCTTTT TTGAGCGTTG CCTTGGCCTC	8160
TGCAACTGTC TGACCTGCCA CATCTGGAAT GGCAATGGT GCAGGAGTTC TGGATAGTAT	8220
CCAAATAAGA GAAGCTGCCA CCAATACAAG GCTGGCCAAC AAAATCAGGT AACGCATCTT	8280
AAATCTATGT TTTTTCGGTG CTGTGTGGTG GTAAGTTTCC TCTGTCACAG CCTGGCTTGG	8340
GTTTTTGATT GATTTGTGTT CTGTTTGCGC TTGAACCTTA GGAATAGATG TCAAGGTAAT	8400
CTGAGAAACC TTCGGCAAGG TCTTGGTATC TGCCTTGCTC GTTTCATCAA AGATTAACTT	8460
ACTTTTATTT CTACGATTGT AGGACAAGCT ACTAGACAAG TCCACATACA TCTCTGAAAC	8520
CGAGCGGTAG CGATTGGTCA ACTTTTTAGC AGTTGCCTTG ATAATAACAT TTTCTAAAGC	8580
CTGAGGTACA GATGGATTTT CTGCAATAAC GGACGGCAGG GGTTCCTGGA AATGCTGGAG	8640
GGCAATGGTC ACCCGCTAT CCCCCTCATA AGGGATATGG CCTGTCAGCA TCTCATAGAA	8700
AATAATCCCC ATGGCATAGA TATCACTCTG CACAGTCGCC TTCGAACCAC GCGCCTGCTC	8760
TGGTGACAAG TAATGAACTG AGCCCAACAT CGAGTTAGTC TGGGTCAGAC TTGTCTCTGC	8820
AAAGGCTACA GCAATCCCAA AGTCTGTGAC CTTGGCAGTC CCATCTGGTG TCAAGAGGAT	8880
ATTTTGAGGT TTCAAGTCCC TGTGAACAAT TCCTCGAGTA TGGGCCAAGC GCATAGCCAA	8940
GAGAATTTGT CCCATGATAC GGACTGCTTC TTCAATTAGAA AGAGGATAAT GTTCCTTGAT	9000
ATAGCGTTTG AGGTCCAGTC CAGCCACATA CTCCATAGCT AGGTACTGTT GACCGTCTTC	9060
CTCGCCAATA TCTGTTATCC GAACGATATG AGGATGGTCT AGATCTGCCA TAGCTCTCGC	9120
TTACGCTGA AAACGAGCTA CAGCTATCGG GTCCGCTCGG TAGTTGGTCC TCAGAACCTT	9180
CACTGCCACT TCTTCCCCAT CTAAGATTAA GTCTTTGGCT AGGTAGACAT CCGCCATACC	9240
TCCTCGACCA ATCTGTTTGA CAATCCGATA GCGTCCGGCA AAAATCTTGC CGATTGGAT	9300



699

CATTCTGCAT CCTCCTCGTT CATAGAAACA AGGGCAACCG TAATGTTGTC TAAACCTCCT	9360
GCATTGTTAG CAAAACGAAC AAGTGTCTCC GTTTTATCTG CTAAAGGAAT ATCACTGGTT	9420
ACAAATATCAC GAATCTCACT GCCTGAAATC ATGTTGGTCA AGCCGTCCTT ATTGAGCAAG	9480
AGATAGTCAC CTGACTCAAG GATAACTGTC CCAAATCAG GCTGAATTTC ATCTTTTTCG	9540
CCAATAGACT GGGTGATAAT ATTTTTTTCG GGATGAGCTT CTGCCTCTTC TGGTGTCAAT	9600
-----	
TGACCAGCCT TGAGCAATTC ATTAACCAAG GAATGATCGC TCGTCAACTG ATGGTATTCT	9660
TCTCCACGAA TCAAGCCGAT ACGCGAATCA CCAATATGAG CATAGATAGC CTGATTATCA	9720
ATAATAGCAA GGACTTCCAA AGTAGTTCCC ATGCCTCTGT AAGCTTCATC CTGACCAAGC	9780
TGGTGAATCT TTTGATTTTC AATTCTAGG TAATGGGCGA ACCATTACAG CACTTCATTG	9840
ACTGTATCGA TCTGGGTATC AACCCAAGCT ACACCCAGGT CTGTGACCGC CATTCTACTA	9900
GCGATATTCC CTGCGCGATG ACCTCCCATC CCATCAGCTA AAATAATCAT GGTACGTCCA	9960
GCTCTATTGA CATAGTGGTT GACATAGTCT TGGTTATTTG TTCGTTTCTG ACCAACATCT	10020
GTTAATAATG AAATTTCCAT GTGTCAGTTC CTTCTAATC CGATATCTTG CGAAATTGAC	10080
TGATGAAGAA TCCATCACTT CCATACAATT CAGGTGTAAT GAGGATACAG CCGTCTTTCA	10140
TGATATCCTT ACATTTCATG TCTAGTTTTA CCTGCTCGAA CTCGGGATGA CTCTCTAAAA	10200
AGGCCTTAAC GACTTGAAAA TTCTCCTCTG AGACGATAGT GCAGGTGCTA TAAGTTATTA	10260
TACCACCTTT GCCTAGTATT TGACAAACAC TACCTAATAT TTCTAACTGA ATTTCTTGCA	10320
AGGACGCGAA ATCTGCCGTT TCTTTATTGT ATTTGATATC TGGTTTTTCG CGCAAAAGAC	10380
CGATTCTGA ACAAGGAGCA TCCACCAAAA TCTTATCAAA GGAATCCTGG TCAAAAAACT	10440
CATGCACCTT TCTGGCATCC AATTTTTGAG TTTGAACCCG ATCTTCAACT CCCAGACGTT	10500
GGGCATTTTC TTGAATTAAA TCCAACTTGT GGTGCTACAA GTCCAGAGCA GTAACCTGAC	10560
CTGTCGTAAG ATAAGAGGCT ATATGGGCTG TTTTCCCACC TGGAGCCGCA CAGGCATCAA	10620
GCACTCGCTC ATCACCTTGT AAATCAAGCG TCGGAGCAAC CAGCTGACTG GACTCATCTT	10680
GGATGGTAAT GGCTCCATCC GCAAACAAAT TATGCCCTGC AAAATGCCCC TGCTCCTTAA	10740
CCAGACCACT GGTGCTAAA AGGGAATTAT TCGCCTCCAA CAAGGCTTGG ATTTCTCTT	10800
TTGACTTAG GTCTGTTACA CGAATACTGG CTTTGTTCG CACTAACAGG CTTTCAAAGA	10860
TGGCTTTTTC TCTCTCCTCT CCGTATTCTT CCTTGAGTTT GGCAACTAGC CAACTGGGA	10920
GAGAATAGGC AATGGAGTCA CGCTTGTTTT TTCGCTTGAT GCTAGCAATA TCTGGCCAGC	10980
CTTCACGCAA GATACGGCGA AGGACAGCGT TGACCAATTT TTTTACGGA	11040

700

GTTTGGCCAA TTCCACTGCT TCATTAACCA CAGCATGATC TGGAACTCTG TCCAAATAGC	11100
GGAGTTGGTA GGCACCTCATG AGAAGAAGGA CATAGAGCCA GCTGTCTAAC TGGTCTCTGT	11160
CTTCGATAAA GTGGGATAGG TACCATTCCA GAGTCAGTTT ACGGGCTACC GTTCCATAGA	11220
CCAGCTCGGT CACTAAGCCC TTGTCTGCTG CAAAAGTTG ACTTCCCTTT AGATGCTTAT	11280
TTAAGGCGAT ATTTGAATAT GCTTGCTTCA CAAAACATC CTCTAGCACT GCTAGAGCTA	11340
AACTTCTAGC CGTTTCTACT TTAGTCACCA AATCGTTCTC CTACAGTCAA TGTACGTCCA	11400
ACTCCGTTGA GGAAGGAAGC AATGTCCATC TTAGGCTTAC CAGCTGGCTG CACTTGTTTG	11460
AGGGATAGAG CCCCTTCAGC CGTTGCGACA ATCAATTCTT TCTTGCCGAT AGAGAGAATC	11520
TCACCTGGAT TTCCCTGACC TTCTACTGGT AGGGCTTCAT AAATCTTAA GCGGTCGCCC	11580
TTAAGGAAAG TATGGGCAAC AGGCCAGGGG TTCATTCCAC GAATTTGGTT AAAGAGTTGA	11640
CGATTGGTTT TGTTCAGTC CAGTTTTTCT TCCTCTGGCT TTATATTTGG AGAGAAGGTA	11700
ACCTGACTCG TATCCTGCGG TTCAGGTTTG ATATCACCAG CAATATAGGC AGGCAGAGTG	11760
TCCAAAAGCA AATCACGACC AACTAGCGCC AATTTTCAA ACAAGGTGCC AACATTGTCC	11820
TCATCTGTGA TCGGAATGCT GCGACGAGAA ATCATATCTC CTGCATCCAT TTCCTTAACC	11880
ATTTCCATGA TGGTCACACC AGCTTCCTCA TCCCCTTGAA TCAAGGCATA ATGGATAGGC	11940
GCACCACCAC GGTGTCTAGG AAGGAGGGAG GCATGAACGT TGACAGCAAA GTCCATGCTA	12000
TCAAGGAGTT TGCTTGGGAG AAAGTCCCCA AAAGCAGCAG TCACAATTCC ATCTGCTCCT	12060
AGCTTCATAA GATCTTCCAT CTCTGGACTT CCAGATTAAT TTTCAGGTTG GTAGATAGAT	12120
AGTCCTGCTT CCTTGGCAGC CTGCTTGA CTGGGTTTCTT GGATAACTTT TTTACGACCA	12180
ACAGCACGGT CTGGCTGGGT CACAACGGCT AGAATTTCTG AACGGTCATC TGTCAAAAGT	12240
CCTTTTAAGA CTGTTGCTGA AAAGTCGGGG GTCCCCATAA AGATTAGTTT TGTCAATCT	12300
TCTCCTTCTT ATAAAAATTG CTGCGGCTCA TGGTCAATCC TGAGACGGAG CTCACTATTT	12360
TCCCGTTCTT GAGTCAAGGC TAAACCTGG TTGAGGGTGG ACCCCAGCTC ATCTTCTAAA	12420
CGGTATTTAA TTAAATCTG GTAATGATAG AGGTTGTGGG TACGGGCAAT CGGTTTGGC	12480
GTTGGCCCCA GAATGGGACT GGTCTCTGAC AAGCCTGACC GCAAAATGTT CATGACTTCA	12540
TAGGCACGTT TGAACCTC TTCTCTTTC TTGTGAGAAA GGGTAATACC AATCGTGAAA	12600
TAGTAAGGTG GATAGCCGAG TTGTCGTCTG ATTCCCATT CATAGGCATA AAAGCCTTCG	12660
TAATCTTGAT CCTTGGCAAA TCGAATAGCA TAGTGCTGCG GATTGTAGGA CTGTATCAAG	12720
ACTTGACCTG CCTTTTCAGC CCGACCTGAT CGACCTGCCA CCTGAGTCAA GAGCTGGAAG	12780
GTTCTCTCAG AAGAACGGAA ATCAGGCAGA TTCAAGGCCG TATCCGCATT TAGAACTCCG	12840

701

ACTAGGGTAA CATTTGGGAAA ATCCAAACCC TTTGCAATCA TCTGAGTACC AAGTAAAATA	12900
TCCGCTTCCC CTCGCCCAA CTGGTCAAGC AAGGCTTGGT GACTGCCTTT CTTTCGAGTC	12960
GTATCCACAT CCATCCTCAG AATGCGAGCT TGGGGAAAGA GTTCTGCTAG CTCATCATAA	13020
GCCTTCTGAG TTCCCGTCCC ATAGTAACGA ATACTGCGGC TCTTACAGTT AGGACAGACC	13080
TGAGGAATAT CCTTCGAGAA ACCACAATAA TGGCAGTTCA TAGTCTTGGT ATCCATATGC	13140
AAGGTCAGAG AAATATCGCA GTTGGGACAA GTATCCACCG TCCCACACTC CCGACACATG	13200
ACAAAGCTAG AATAACCACG GCCATTGAGC ATGAGAACCA CCTGCTCTTT TTTAACCAGA	13260
CGGTCTTGGA TAGCCTCTAG CAAAGGAGGC GTAAAGTTTG ACGTCTCATT TTGTCCGATA	13320
TAGTCTCGAA AGTCAATCAC TTGAACCTCA GGGATTGTAG CCAAAGGATT GGCACGTTGG	13380
GTTAGACGTA AGTGTTGATA GACGCCTTTG CCAGCACGTG CCCGGCTCTC TAAGCTCGGC	13440
GTTGCAGATC CAAGTACCAG AGTTGCTTGA TTATACTGAG CCCGTAAAAT AGCTACCTCT	13500
CTGGCATGGT AACGGGGATT GCTGTCTGCG TTATAAGCCG CTTTCATGCTC TTCATCAATA	13560
ATCATGACAC CCAGATTTTT CAGAGGAGCA AAGATAGCAG ATCTGGCACC AACAACAACT	13620
TGGGCATCGC CACGCTCCAC CTTGCGCCAT TCATCATACT TTTACCAATT GGATAATCCT	13680
GAGTGAAGAA TGGCTACCTT GTCCCCAAAA CGTGCTATAA AACGCTCGGT CATCTGAGGA	13740
GTCAAGGAAA TCTCAGGTAC CAGCAAAATA GCTGTCTTGC CCTTATCCAG GGCACCTTGG	13800
ATAATCTGCA AGTAAACCTC GGTCTTCCCA CTTCTGTAA TCCCTTGAAG TAGAAAGGGA	13860
GGTTGAGAAC TGCCAATACA ACTCACAACC GCATCACGGC CCTGTCTTTG TTCTGGATTT	13920
AACTCCAAAG GTCTACTTGC TTCAATTCCT TCAAAATAAG CAGCCGAGCG TTGAACTTCC	13980
TTTTGGACTA TGGTAACAGC ACCTTGATCC ACAAAGAAGT TGACTTGCTC TCGCGAGTAG	14040
GACTCTAACA AGCTAGCCAA GGAAGCGCTC TCTGGATGAG ACAGCAGATA ATCTCTCAGT	14100
TCCAACTTTT TCTTGGCAGC TGTAGAAATC TCAACACCTT CTAATTGAGC ATGGTCAACC	14160
TCATACCAAG ACTGGGTCTT GACCTTCTTT TGATCGACTC CCTGATATTC CAGACCAAGC	14220
AGGCCTTTTC TAGTCAAACG CATCATTTCA GCTTGCTTGG CAAGGTCTAG TGAAGAAAAG	14280
GCTAGCGAAT CTTCTGAACC AAACAGGCGC ACTCGTTCTT CCTGACTCAA GCCTTCGAGA	14340
GGATAGAGAA TCTTGTCATA GCTAGAATTC AGAAACCCTG GAAGCATGGC CTTGAGGATA	14400
GAGATTTTGT AGGAGAAGAC AGATTTGCGT AACTCCTCAG CCAGCCAGAG TTGTTCTGGC	14460
GTGAGAACAG GAGAAAAATC CAGCACCTCT GCAATATCTT TTAATCTTG CTCCATCTCT	14520
TCTCCATCTG ATTGGGACTT CAAACCAAGA ACAATCCCTT GAATCAGGCG ATTACCCCTA	14580

702

CCAAAAGGCA CATGAACCCG CATCCCAACT TCCAGCATTC CCTCAAATTC CTCCGGAATC	14640
CTGTAAGTAT AGGGCTGGTC CGTCTGCATC AAGGGCACAT CTACGATAAT CTTAGCTAGG	14700
GCCATCTTCT CACCTCCTCC TTGTCAGTAC ATTCTTGCAA TAGAAAAAAT AAGATTGAGT	14760
CCCCCAACC TTAAATTTTT TCACCATCTT CTTTTCTTT AGCAATTTGC TCTTTGATTT	14820
TCTTTCTTC TTCTTCTTG CGGCGTTTTT CTCTTCGAT ACGGCGACGC ACTGCTTCAC	14880
GTITTCCTTC TGGATCTGGG TGAATTGTAA CGTTTCCTGA TTCGATTCTT TCTAAAGCGC	14940
GAAGAGTTGA TTTTCAGAC TTGAAACCTT GAGTTGCTGG GGCACCTGCT TCCAATTCGT	15000
GGGCACGTTT TGCTTCCAAG ATTACGAGTG AATATTTTGA AGGAACCTTG TCGAGCAAGG	15060
TATCAATAGA GGGTTTTAAC ATCATTTGCT TGTACCTATT TTCTAAATTT TATCGGGTAG	15120
TTGGAGATTT TGGTAACATC TCCTGATAGT GACCAATGAC ACGATCCACA CAGAAGTGTT	15180
CTGCTTCAAT CACACATTG ACACGTTTAC CAGCTAGGGG TACCTGATCG TTGACAATCG	15240
CATAATCATA CTCACGCATG AGGGCAATTT CTTCCTTGGC CTTTTCGATT CGTTGGGCAA	15300
TCACTCTGC ACTATCTGTT CCACGACCTA CCAAGCGATC TTGCAATTCA TCCAAATCTG	15360
GTGGTGTGAG GAAGATAAAG ACAGCATCTG GAACCTTTTT CTGACCTGA AGAGCACCTT	15420
GAACCTCAAT TTCAAGGAAA ACATCGATTC CTTTGTCCAA GGTTCATTG ACATAGGTCA	15480
GAGGAGTTCC ATAGTAGTTA CCGACATATT TGCGTATTC CAACATCTGT CTTGACGAA	15540
TCAGCTCTTC AAATCTTCA CGAGTACGGA AGAAATAGTC AACACCGTCC ACTTCTCCAG	15600
GACGTGTGTC GCGTGTGTC ATCGATACAG AATATTGAAA TTGGTTTTCA GAACTCTCAA	15660
AAATCTCTCT TCTAACCGTT CTTTTCCAA CCCCTGAAGG ACCAGAAAAA ACGATTAGTA	15720
AGCCTCGGTC TGCCATTGTG TCTCCTTTTA GTCAATCTGT GAAATAACAT TTCTCTAGAA	15780
TAATGGCAAA AAGCCAGATT ATCCTTTACA GTCTTTCTAT CTAGTGTAAC AAAAAAGCAG	15840
TAATTTTTCA ACTGCTCTTT CTTATTTATT TAGCATAATC TACTGCACGA AGCTCGCGAA	15900
TCACGGTTAC CTTGATATTT CCTGGATAAT CGAGATTGTT TTCAATTTTC TTACGAACTT	15960
TGTGAGCCAA GATTGTGACT TTGTCGTCCT TGATTTTTCC TGGATTGACC ATGATACGAA	16020
TTTCACGTCC TGCTTGAAGG GCAAAGCTAG TTTGCACTCC TTCAAAGCCG TTAGCAATTT	16080
CTTCCAAATC ATGGAGACGC TTGATGTAGC TTTCAAGAGA CTCACTACGA GCACCTGGAC	16140
GGGCTGCGCT CAAGGCATCT GCTGCAGCGA CGATAACTGC TATCACGCTC TCAGCTTCAA	16200
CATCTCCGTG GTGACTAGCA ATCGTATTCA CCACAACCTG GGGTTCCTTG TACTTACGGG	16260
CCAATTCAT ACCGATTTCA ACGTGGCTAC CTTCAACCTC ATGGTCAATG GCTTTCCCGA	16320
TATCGTGAAG GAATCCAGCA CGACGGGCAA GAGCCGCATT TTCACCAAGT TCGCTCGCCA	16380

703

TGATACCAGC	CAACTTAGCA	ACCTCAATCG	AATGGCGCAA	AACATTTTGT	CCATATGAAG	16440
TACGGAACTG	CAAACGTCCC	ATAATCTTCA	TCAAGTCTGG	ATGAAGGTTT	GGCGCACCAA	16500
TTTCATAGGC	AGCAGCCTCA	CCGTATTCAC	GAATCTTATT	GTCAATCTCT	TGACGGTTTT	16560
TCTCAACCAA	CTCTTCGATA	CGAGCTGGAT	GTATACGACC	ATCTTTGAGC	AACATTTCCA	16620
TAGTCATACG	GGCAATCTCA	CGACGAATCG	GATCAAATCC	TGACAAGGTC	ACCACTTCTG	16680
GTGTATCGTC	GATAATCACA	TCGACCCCTG	TCAAACTTTC	AAAGGTACGA	ATGTTACGAC	16740
CTTCACGACC	AATAATGCGT	CCCTTCATAG	TATCGTCTGG	CAGATGAAGT	GTTGAGTTTG	16800
TTGACTCCGC	TACATATTCA	CCAGCGATAC	GTTGCATAGC	TTGAACCAAG	ATGTCCTTGG	16860
CCATTTTGTC	AGAACGTTCC	TTGACCTCTT	GCTCAGCTTC	GCGAATGCGA	CTGGCAATCT	16920
CCCTGGTCAA	GTTTTCTCT	GTCTGAGCCA	AGATAATATC	TCGTGCTTCT	GCCTGAGACA	16980
GCGCACCAAT	ACGCTCTAGT	TCTGCTTCTT	TTTGTCTTTC	GACTTCCTCT	AATTGCTCTT	17040
CACGCGCATC	AAGGTTTTTC	GCTCTATCAG	AAATACTTTG	TTCTTTTGT	TCAAGTGTTT	17100
GTTCTTTACT	CGTCAAATTG	TCGTCCTTAC	GGTCAAGGCT	AGTAGCTCTC	TCTGTCAAAC	17160
GACTTTCGAT	TTGTTTGAGT	TCTTGACGTT	CTGATTTGAA	TTGAGCGTCC	ACTTCTTCAC	17220
GGTATTTTCT	GGCTTCTTCT	TTGGCCTCCA	ATAGTGCTTC	TTTTTTAAGA	GACTTGCTTT	17280
CACGTTTGGC	TTCAATTAACA	AGTAAATCCG	CTTCACGCTC	AGCTTGTCCT	CGTAAATTAG	17340
TTGCTTCTTG	TTGAGCATTT	AAAAGCATCA	ACTCTGCAGC	TTCTTGAGAT	GATTTTCATCT	17400
TAGCTGAGAT	GCTGACATAT	CCAATGACTA	AACCAATGAT	GACGGCAAAA	ACAGCAATCG	17460
CAAGCGACAT	GATTTCCATG	TTTTTACCTC	ATTTTATTGT	TATTCCTGAAT	GACATACATT	17520
CTTTTACATT	CTACCATAAA	AAAGTGATTT	TCACAAACCT	AAAATAGAAT	ATGTTTTGAG	17580
GAATTTGGAA	CACATTTACC	AAAATAAACT	TGTTGTTTAG	AAATAGTAGT	TTAGTAGAGA	17640
CTTGAGAAAA	AGCCTACCTT	TCAATAGACT	TAGTAATGAT	CTTTAAAGGA	CAAGAAAGCC	17700
ACGCTATCTC	CATCCATCAT	ATAAATCAAG	CGATTTTCTG	CATCAATACG	CCGTGACCAG	17760
GCTCCTTGGT	AATCATATTT	GAGTGGTTCT	GGTTTACCTA	TTCTGTAAAA	GGGATCACGT	17820
TGAATATCCT	TGATTAGTTT	ATTGATTCTT	TTTAACGTTT	TCTTATCCTG	ATTTTGCCAG	17880
TAGCAATAAT	CTGCCCAGGC	ATCTTCTGTA	AACTTGAGCA	GCATTTCTTA	CTCCTCAATA	17940
ACATGGACCT	GAGTACTTCC	AGCACGAACT	TGAGCCATTC	CTCGCAAAAC	CTTATCAGAA	18000
AGTTCCTTAT	TTTGAGCAAT	TCTCAGGGTT	TCTTGGATAC	TATCCCACTC	ACTCTTTGAA	18060
AGGACTACAA	TGTCCTCATC	TGGATTTTTA	TTGACCACCG	TCAAAGGCTC	AAATTCATCA	18120

704

TTTACCTTCT TCATGTAGTC CTTTAAATGA TTTCGGGAATG TTGAGTAAAG GACTGCTTCC	18190
ATAACCATAC CTCGTTTTC TAGCTTTTCCA CTATTATACA CGAAAAGAAA GAAATTGTCA	18240
GGAACCTTGT CAAGATTTTC TTTTCTATCT ATTTATACTC AATGAAAATC AAAGAGCAAA	18300
CTAGGAAACT AGCCGCAGGC TGTACTTGAG TACGGCAAGG CGACGTTGAC GCGATTTGAA	18360
TTTGATTTTC GAAGAGTATT ATTCGTAAAA AATCTCAAAA AGCCTACCTT TCGGTAGACT	18420
TAGTTTGTTC CTATTC	18436

(2) INFORMATION FOR SEQ ID NO: 88:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7001 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ACGTAGAAAA ACTATTTCTA TCACAGATAA TATCCCGTAT GTTGTGGAG GTATTGAAAT	60
AAACGTCCTA GGTATCTTTC TCAGTCTATG TGACTTACAA GGGAAAATC TTTCGAGAC	120
AGAAATTTTG AATGAAGATT ATCCTATTTT AGAAATCAAT TCCACCATT CCAATATGAT	180
AAAAACAGCT ATAGAGTACG TCCCTTTGGA AACAAAATTA CTTGGATTG GCTTATCAAT	240
ACCTGGACAT TATAACAAAG ACTCCGGAAG TATCATTACA AACAAACCCA TATGGGAATC	300
TTTAAATTTA TTAAATGTAA TTTAAAGATT CAATTTTCCT TTTATTGTAA AAAATAATAT	360
CGATTGTATG GCTATAGGAC AATACCTTTT TAATCCACAC AATACCCCG ATAACCTTAT	420
TTTCCTACAC GCTGGATTAG GTATTTACAC TTCCTTTTTC ACAAAGAAA AAATAGGAGC	480
CTCTAAAAAT CCTTATATCG GAGAAATTGG ACACACCATT GTCGAATTGA ATGGGCAATA	540
TTGTGAATGC GGAAAAAAG GTTGTTTACA AACATATATT TCGGATGCTT GGTTAATCAA	600
ACACGCCCAA TTATTATTTA AAAATTCCCA ACTAACTGTA CTAAAAAGCC TTGTAAAGAC	660
TGAAAAAGAC ATTCATTAG ACACCCTTTT AACGGCTTAT AATTTAGGCG ACTCCGCTTT	720
ACGTCAACAA ATTGATAAAG GAGTCAATTT ATTAGCCACT TCTATTGCAA ATCTCCTCCT	780
CATCAATCCT GCTGATAAAA TCTATATCAA CAGTCAATTG CTTAATTATC AACCTTTCAC	840
TCATGAAGTC AGGGATAAAA TCCAAGACCA GCTCCACTTC GTTCCCTTTA CTCGTAATAT	900
AGAAATTGAA ATTTTACCTT ACAACAAACA TCGTGGAAGT ATAGGAGCTT GTGCATTAGC	960
TATCGTCGCT TTTTTCATAG AACATAGCAA TGTATTACAA GATATTATTT CACCTTAATA	1020
TATTAGAAAT CTATAGACCT GTTTAAATCA ACTATAACCT GTAGTAGATA TCTCGTATTT	1080

AGACAATATG AAAACAAGAC GACTTCCATA TAGGAAACCG CCTTCTCGCT ATGTTGAGTG	1140
ATTTATATTA AAATAACTTT TCTTCTAGCT GCATTTTATT ATTATAAAAA CATTTCATCAT	1200
AACCCCCAGA ACTTAAATAA CAATTTTAT TCAAGATACA TACTCCTAGA ATAAACTTTA	1260
TATGAAATTC TCATTTTGT TTTTACAATT CTCCTTAGTT AAATCTTGT TAATATATGT	1320
TTTACATATA GTATTTAGCG CCACATAGTA CTGAACTCTC TCCAAAAACG GTTATTCCTC	1380
TTTGAATAGG GCGTTATCAC AAGAAAAGCA TCTCCACGTT TCAACTTCAT ATGGCTCAAA	1440
AACAATCAAT TGATGCTAAA ACCTGTACCT AGATGTTTCG GTTCATAAAA CCATGAAACT	1500
GTAAAAGTGG ATGAAATGA TAGCGATAGT CAAATCAAGA GGCATCATAA CTCTAAAAAG	1560
TCACAATATA TAAGTTCATC CTCGGAAAAA TATCATTCTA ATTGTTGAAA TGCCTACATG	1620
AAAAGAAACG TCAAATGCTC ATGAAACAAC GAATACAGGT ATCAAACTA TGACAAAACA	1680
AATCCCTAAA TTTACTAAG AACTGTCTCA ACTTTACACC TGTAATGGT TGTGTATAA	1740
TAAAGTTACA AAGATGTACG ACCACACTGT TGTAATCAT AGTGTTCCGG AATATATTAC	1800
TGATAGCATT TCTACAAATA CAAGTAAAGA GAGCGGATGA GATTCAAACG AAATATGTCA	1860
GTGCTTTGGC ATTCCTAGCC TTCATATCAT TTAAAGAATT CTATAGACAA AATTTTTTCC	1920
AATACAGACA CTCGTAACAA CTGCTTCATT TTTCTACCA CATATTTAGG AACAGGATAA	1980
GATACAAGAG TATTAATCCA TAGCTCAGTT CTATACCAAT CTAAGACAAA TAAGCTAAAA	2040
AAACGATTGA TAATAAGCAA ATAGATTCCA AATTTTCTCT ATCTGCTCAT TTTAATAAAC	2100
AATACTAGTG TAACTATCCT TCCAGTCAGA AGCTTGTCAG ATCACACCGA AAATCTTCT	2160
AAAATTTATC TCGTTAGGCA ATCAAGCAAA AACTCGACGA TAGTACAAAC ATTATCATAC	2220
AGGATTGACT TCCTAAATTA TATACTTTAG TAAGGTTTTT GGATAAGAAA AAAGGTTTCT	2280
TTTACATTTT TAAACATTCT TTTCTAAGAT GAAAAACAGA ATTTTTCGAT TGTGATTTAA	2340
AGCAACAAGA AGATTTTCAG TATCATCCTA TAGATACGAG CTAATTAAGA AAACTACAT	2400
TTTTGAATAT AAACACAAT AATATAAACT AAATTTTATA GGAGGAAGAC AATGGATTGG	2460
TACGATTATA TGATACAGGC ATCCAAACAA TCACAATTCA ACGCAAGCCA TTGGTTTCGC	2520
TATTTGCCAA AAGTTATTTT TGAAGACTAT TCTTATTTAA CAAACCAAGA TGTAGAAAAG	2580
TTGCTAGACT CCAAAGAACT AACCCGTTTT CAAAAATTA GCTTGAAGTA TGCCTTTCAA	2640
GAGCATACTC CAACTCATAA ATATGTGATT TCATTAAATA AACCTGCTAA GTTAACCAAT	2700
GTTCAAAAAT TGATGGAGAA ATACAAACAT GGATAAAATG AAACCGGTCT TCCAAGCCCT	2760
AAATAAGGAA TTAATTCAGG AAAATCTGAC TTAAACAATT ATCTGTGTCG GTGGTTATGT	2820

706

CTTAGAATAT CATGGTTTAC GTGCCACACA AGATGTTGAT GCTTTTATGG CTCTATAATA	2880
TTTGTAGTGG GTAAATCCCC TATGGATATT ATGGAGCCTA TTTTGTGTA GAAAAAAGT	2940
CCCATATGAC CTATAATGAA AAGCGACAAA ACAACTCATT AGAAAGAATC ATATGGAACA	3000
ATTACATTTT ATCACAAAAT TACTAGACAT TAAAGACCCT AATATCCAGA TTTTAGACAT	3060
CATCAATAAG GATACACACA AGGAAATCAT CGCCAACTG GACTACGACG CCCCATCTTG	3120
CCCTGAGTGC GGAAACCAAT TGAAGAAATA TGACTTTCOA AAACCGTCTA AGATCCCTTA	3180
CCTCGAAACA ACTGGTATGC CTTCTAGAAT TCTCCTTAGA AAACGCCGTT TCAAGTGCTA	3240
TCACTGTTCA AAAATGATGG TCGCTGAAAC TTCTATCGTC AAGAAGAATC ATCAAATTCC	3300
TCGTATTATC AACCAAAAAA TTGCGCAAAA GTTGATTGAG AAGATTTCTA TGACCGATAT	3360
TGCTCATCAG CTGGCCATTT CAACTTCAAC TGTCATTGCG AAGCTCAATG ATTCTCACTT	3420
TGAGCATGAT TTTTCGCGTC TTCCTGAGAT TATGTCCTGG GACGTTGAAA CAGTCCGGGG	3480
AGTGACTGTT TCAATCGGGA GATGGAGATG AGCTTTATTG CGCAAGATTT TGAAAAGCTC	3540
GATATCATCA CTGTTCTTGA AGGTAGAACA CAAGCTGTCA TCCGAGATCA CTTTCTTAAA	3600
TATGATAGAG CCGTCCGATG TCGCGTCAAA ATTATTACTA TGGATATGTT TAGTCCTTAC	3660
TATGACTTAG CTAGACAACT TTTCCCGTGT GCTAAATCG TTCTTGATCG CTTTCACATT	3720
GTACAACATC TTAGCCGTGC TATGAGTCGT GTGCGTGTC AAATCATGAA TCAGTTTCAT	3780
CGAAAATCCC ATGAATACAA GGCTATCAAG CGCTACTGGA AACTCATTCA ACAGGATAGC	3840
CGTAAACTCA GCGATAAACA TTTTATCGC CCTACTTTTC GTATGCATTT AACCAATAAA	3900
GAGATTTTAG ACAAGCTTTT GAGCTATTCA CAAGACTTGA AACATCACTA TCAGCTCTAT	3960
CAACTCTTGC TGTTCACCTT TCAGAATAAG GAACCGGAGA AATTTTTCGA ACTTATCGAG	4020
GACAATCTTA AGCAGGTTCA TCCTATTTTT CAGACTGTCT TTA AACCTT CCTCAAAGAT	4080
AAAGAAAAGG TTATCAACGC CTTTCAACTA CACTATTCTA ATGCCAACT GGAAGCGACC	4140
AATAATCTCA TCAAACCTAT CAAGCGCAAT GCCTTTGGTT TTCGAACTT TGAAAACCTC	4200
AAAAACGGA TTTTATCGC TCTGAATATC AAAAAAGAAA GGACAAAATT TGTCCTTTCT	4260
CGAGCTTAGC TTTTTCCTT CCACTACAG TTGACAAAGA GCCGGAAAAA GGAACAGCCT	4320
TAGCTTTCCT TTCATTCTT TTTATTTCCC TCGTAGTAAA CGTGCTAGCT TCCACAAAAC	4380
AAACAGGATT CCCAGAAATG CCAGTACCAC TAGCCACGG TACAACCATT GAGAGGTTGC	4440
AACACGCGAT ACAGATTGTC CTTCTTTCGT AAAAGCAACC CTCGCACTG CAGCTGTTTG	4500
TGGATCTGAT TTTTGATAAA CAGCGACTCG TTCAAAATTC ACTAATAAGC GTTTATTAAA	4560
GGTAGGAATC GGATCGCAGG TTATCAAGGT CATGATATTT TTAGAGCTAA CCGATTCTAA	4620



707

TTTTTCCCAT	TCCGACGGTA	AAATAATCTC	TGTGTCCATC	ATCTGATATT	CTACAATTTT	4680
CTGGCCATTA	TCATAATAAA	GAGCATCTCC	AACTTTTAGC	TGATCCAAAT	GGCGGAAAAA	4740
GACATGGCTT	GGCTCTGCAC	GGTGCCGAGC	AATCACTGAG	CGAATCCCTG	TACCATCCAG	4800
AGGCAGCGGT	GTACCATCCA	CATGAGCCAA	GCCCATCCCT	AAATGATGAT	AATCTGCTCC	4860
CAAATAAAACC	GGCTCCATGA	TTTCCAAACT	TGGAATAGAC	AAGTAACCAT	AGACTGCATC	4920
AGGGTCGTCA	GACACTTGGT	AATTGACCTC	ATATCCCTCC	GCCAAAAAAG	GATCTACAAT	4980
GCGATTTTGC	GAAGCCAAGC	GTTGATTGTA	GGCGAGAGAA	TGGTTCTGTT	GTTCTTGTTA	5040
CATTTCAAGT	GTCATGGATT	TCACAAATGT	AGCATGACCT	TTCACCTGTC	CAAGAGACTG	5100
CAACACCATC	TGTCCAAAAC	AATAAATAGG	AATCAAACAG	GCTACCAACA	TCAACAAGTA	5160
TCCCAATAAG	GCTCGTAGTT	TAGTCCTTGA	CATGACGCCC	CTCCAATTGC	TTTTCTAGTC	5220
CTTTGACAA	CCGTCGATTA	CGATACACGC	GATACAGCAA	GAGAAGGATG	ACCGCCATCG	5280
CTCCTAGTAA	TAACCACAAC	CAGAAATGCC	CACGCTCTCT	CACCGCTCGA	TTCCGCTCTG	5340
CAATTGGTGC	CGTATACGGA	ATCCGCTTCC	CACGTACCAA	CAGACGATGA	CTGTTAATCA	5400
TATACGGTGT	ACAAGTCAAC	AAGGTCGCAT	AATCTTCCCC	ATGTTGAATC	AAGACAGGCT	5460
CAAAGTCATT	CGGCTCCACC	GTCACTATCT	GATCCACTTG	GTAGGCCAAC	ACCTGATCTA	5520
AAACGTGAAG	ATAAAAGATA	TCCCCTTTTT	TCATCTTATC	CAATTGACTG	AACAATTCTG	5580
CCGTGGCAA	TCCTCTGTGA	GCAGTGATCA	CTGTATGGGT	ATTTTCACCT	CCAACAGGCA	5640
GCGAAGCCCC	TTCTAACAGC	CCTGCCCCCT	TCTGAAGAAT	GTCCTCACTC	GTTCCGACAT	5700
ACATCGGAAT	TTCTGATCA	ATCGCAGGAA	TTTCCACATA	GCCAATCCGC	TCATGGACCT	5760
TTAGCATATT	GGCATATTCT	GAGACGCCTT	TCTTTTTCTC	TTGCTCTGTA	AAAGGATCAA	5820
GAATTTCAGA	TGGTTTCAAG	GTCGCATTGA	AGGCTTGAGC	CAAGCGCCAA	CGCTCCTCAA	5880
GTTCTGCCTT	ATCCATCTGG	GAAACCGTCT	CATCAAACCT	TTTAATAACC	TCGTTTGACT	5940
CAATACGATA	ATAATAACGA	GACACCAATG	GATATATCGC	AACGGCGAAT	CCTACTAAGA	6000
AAATCAGAAG	AAGGATCAGC	GGATGTTTCT	TCTTTTTTGT	GCCTTTTTTT	CGTGAACGTC	6060
TACTGTTGTC	CATCCTCCAC	CTTCACTTCC	TTCTTGCTG	CTTTCAGCGC	CTTCAAAGCC	6120
TTTTCCGGTT	GTTCCTTCTT	CTTGCGCAAG	CGTCGAATAA	TCCATAAAAG	AATCACAATC	6180
AAACCAACTG	CCACATAAAA	CAGGTAGCGA	TAGAGATGAC	TGAGTTTGTT	TGCTGCAATA	6240
AATTCTTCCT	CAACCTCTGC	TACGTACGGT	ATCCGATGCC	CCCGAACCAG	TAGACGATGG	6300
GTATTGATCA	TGTATGGCGT	ACAAGTCAGC	AAGGTCACAT	AATCATGACC	TGGTACAATC	6360

708

AATAAATCAT CAAAGTTCGT CGGCTCAATC ACCTTTACTT GATCCACTTG ATAGGCCATC	6420
ACTTCCTTGA TATTGTGCAC ATAAAACTTA TCCCAACTT TAAGTTTGGT CAAATCCGTA	6480
AACATCTTAG CTGTTGGCAA ACCTGTATGT GCCGTAATCA CCGCATGGGT CGAATTGCCT	6540
CCGATCGGCA GAGAAGTTCC CTCTAGATGC CCAGCCCCTT GCTGCAATAC CTCCTCAGCA	6600
GTACCAGCAT AAACCGGCAA ATCCACGTCA ATAACGGGGA TTTCCACATG CCCCATCCGC	6660
TCATGGATT CTAACATACG TGCATACTCT GCTCGCCCTT TTTCTTCAT TTCTCCGAC	6720
CAAGGATCGC CACTCACTAC ATTATTCAAA GAGTCATTGA AGGCTTGTGC CAATTCATT	6780
CGTTCATCAA TGTGAGCCTC ATCCAACGTT GCTTTTCTT TATCAAAGTC AGCAATTTGT	6840
TGATTTGATT CCACTCGATA ATACAAGCGA GACACCAGCG GATACGCCAT TACCGCCATT	6900
CCAATGAAAA ATACCACTCC TAATAGGAGA TTATTTCTGTT TTTGCTTTT TGTTTTACC	6960
ATTTTATCA GCATCCCTTT ATCTTCAAAC TTCAGGGTAT C	7001

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10411 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GAGGGAGCTT AAGAAGTTAC CACCGTCCTC TAGCGCCTTA TCCGCATCAA AGTTAAGGTT	60
GATATTTTAA AACTGTGCGC CAGCTTGTGA TACGATGCTT TGTTTAAGGT CATTTAGGTT	120
TTTAGTGAAA TCTGCATTGC TGAGGATATC ACTCTTTGAG AGATTCAAGG CAAAATTGAT	180
GATGATATTG ATCTGGTTTC CTGTTATGAC CTGATCAAGT TTGTAATTTT TTAAGGTATC	240
TTCAACAATC TTGCGGATAT CTCTTCTGT CAGATTTCCC TTAATTTCTT TAGCTTTGGC	300
GAGTCCTGAC TTGATATCAG CTAGGGCAAC GTTTAATTTA TTAGCATCAT AGCCTGATTT	360
GTCTTTGTTT TCAGCATTGA TATCTGACAA AGCTTTTAGC TCTTCTTGAG CCAAATCTTT	420
ATTAGCTTGT GGCACCTTGG CTCCATTAGC CTCTAGCGAA TAGTAAATCC CTGCTAAAGC	480
ACTTTCTCCT GTAAGTGGAA TAGGGGCTGC TACAGTGATT TTGGCATGTT CCATACCCAG	540
CGTTACTGCT GCGTTTCGGT ACATATCCTG AGTCACCTTA GTGATATTTT CTGGTGTTC	600
AATCTTGACC TCAAGTGGCG ATTTGTCAAC TAGCTTTTGA ATCTTGGCTG ATGAATACAA	660
CTGTAAGCTA GAGTCATTGG CCACATTCAT GATTTTAGAA TAAACATCAG GTGTCATGGT	720
CTTGAGTTCT TTGGTATCTG TTGAGGCATT GTAGCCCACT TTTTAAAGAG TTTGATTTT	780

709

TTGGTCTTCA GATAGGGAGG AACCTAGGAC ATATTAGGCT TGGACATAGG TTTCATCGAT	840
AACTTTTGA ACATCTGTTG CTGCATGGAC GCTATTGATA GCTGTTACTG CCCACAAGAT	900
CGCAGCGCTA GTCAGAAAGA GTTCTTTCT CATAGGGAAT TTCCTCCTTT ACTTCTTTAG	960
AGTAATATAT CTATCTTAAA GAAACTTAT AACAAAAACA CCTGGTCTAG CCAGATGTTG	1020
AAAAGAGAGT GAAACATTG ATGATGTAAA GGTAAAGTCG TACCTGTCTA GAATAATAAT	1080
AGTTTCCTCC ATTTACATAG AGTTCAGCAC CGTGAAAAAT GGAAATGGGG TGAATATAAC	1140
TATAAGTCTT TCCAGTCCTA TTACCAAGCA AGGGGGCAAC AGTCTCACGA GAGTACTGTT	1200
TGGCTAGAGC CAGGGTATTT TCCTTGCCAT TTGGGGCGAT AAAATCGATA TAGGCAGGTC	1260
CAAAATTATA GGCTTGAACA GCTGTCCAGA TATCTACCCC CTTCTTCTGC GCCAGATAGA	1320
GATTGCCTGT CAGAGTTTGA ATGCCTTGCC GAATGCTAGA GGCATTATCA TTGATGGTGT	1380
TGGTGGAACC ACTTGCAGAC TCACTAGACT GCATAACATC GCCTTCTTTT CTTTTGTTT	1440
CAGTATAAAT CATAGCAAGC ACAAGCTCTT CGTTTGCTGG GGTGTCTTGT TCACTCAATA	1500
TTTCTCGCAC CATGGGTGTA TAGGTCAATG CTTGTTTGAC ATCTTGATGA ACGCGGTAAG	1560
CTTTATAGCC AGCAAAAAGG AAGACTGCTA GTACAAGCAC TCTTCGAATT CGTTTAAACA	1620
TTATTTACTT TGGATATCCT CGATATTTT GATTAAGATA GAGTAGGTTT CATTTTCGTT	1680
TTGGATAAAC TCAACAGACT CGGCGTCTTG ATAGACGTTA TTGGGAACGA TGAGCTCAAT	1740
TCCATTTGAT AAGGAGAGTT TTTGGTTTTT AAATTTCTTT AATTGGCGAC TGGCATCAAT	1800
TTCATCAAAT TGAACAGGTT CTGGTACGGC TTCTTTGACT TGGTCAATAA AGCTCAAACG	1860
AGCCGTCAGA TTGTTGTCAA AAAGGTCATT AGCCAATTTT TCAGGTGACA ATTCATTGCT	1920
TTCTTCTAGG TTGTTGAAAA TAGCTGATTT GACCTTGGAT TGAAATTGAA AATCATCTGT	1980
GTAAAAAGAT TTAGCAATTC TCTGGGCTGT TTTTCCAGT TCCTTGATAG ATTTTTTAGG	2040
AGAAATCTTA GGAGCGACAG CAAGAAGATT ATCTGAAAAA TAGTTCAAAA AAGTCCCGTT	2100
GTACTTGATT CGTTTTTCAA TCAGGTGATA CTTGCTACTC TGAAGATTGA CCACCAAGGC	2160
CTCATCAGCT CCTGTTCCAA ATCCAGGCAG GTTATTCTGA GTTAGCTTGA TTGGATTATC	2220
AACTTCTCCT CCGAGGTGGG TCAAGGTCTC CCGCAGGGCA ATTCGCAAGA AAGCGAAATG	2280
TTCTACACCT TCTTTAGAAA ATTGCACAAA AATCAAGTCA TTGGTCTTGA GATTTTCAGA	2340
AATGCTAAAC TCCTCTTCC AGAGATTAGC CAGCGTACT GATGTCTCCA ACAAATCGTC	2400
TGTAATATGA TTGAAGAAGG GATTTTCTTC TTCGAAAATC CCAGTCTTGG CTTCATCTGA	2460
ATACACATGT TCAATTTTTT TACGCAGGTA TTCTTCGATT TTTGGAGTAA TATTGAGAAA	2520

710

CTTATCTGCT	AAGAACACTT	CGGTATCATC	CGGACTGAAC	TGGTGAATAA	TGGCTTTCTT	2580
AATATAAATG	TCCATAAAAG	TTTTAGTCCT	CGTATAATGG	GAAGGCATCT	GTCAATTCTT	2640
TGACTGCACT	TCTCACTTCT	TCTAATACAG	CCTCATTTTC	TGAATCTTA	AGGGTTTTAA	2700
TGATGAGTTC	AGCCACTTTG	CGACTTTCTT	CTTCACCAAA	TCCACGTGCA	GTAATGGCTG	2760
CTGCTCCGAT	ACGAATCCCA	CTTGTCTTGA	ATGGTGACAA	GCTTTTCGTAA	GGGATTGAGT	2820
TTTTATTAA	GGTAATATTG	ACTTCATCCA	ACAAGTTTGG	AGCAACTTTG	CCGTTTTCTA	2880
CAACTTTAGT	CACATCAACA	AGGAAGAGAT	GGTTTTCACT	TCCACCTGAA	ATAATACGGA	2940
AATCAGGGTC	TTGCAAGAAG	ACATCTGCCA	TAGCCTTGCT	GTTCTTAATT	ACATTGGCAG	3000
CATATTCCTT	GAAGGCTGGA	TCCAAAACCTT	CTTTGAAGGA	AACTGCCTTA	GCCGCCACAA	3060
CATGCTCTAA	AGGACCGCCC	TGAATACCTG	GGAAAATAGC	TGAATTGATT	TTTTTAGCAA	3120
GTTCTTCGTC	ATTGGTCAAA	ATCAAACCAC	CACGAGGTCC	ACGAAGGGTT	TTGTGGGTCG	3180
TTGTTGTTGT	GATATGAGCG	TATGGAAGTG	GGCTTGGATG	AAGGCCAGCC	GCAACCAAGC	3240
CAGCGATATG	GGCCATGTCC	ACCATGAGCT	TCCACCGAC	AGCATCTGCG	ATTTACGGA	3300
ATTTGAAAA	ATCGATAATT	TGAGAATAGG	CTGAAGCACC	AGCTACAATC	AGTTTTGGTT	3360
TTACTTCTTG	GGCTTGTTC	AAGATAGCAT	CAAAGTCTAA	GAGTTCCGTT	TTAGGATCAA	3420
CACTATAAGA	AACAAAGTTG	TAGGTTTGAC	CAGAGAAGCT	AACAGGAGCC	CCATGAGTCA	3480
AATGACCACC	TGATGCCAAA	TCCATTCCCA	TAACCGTATC	ACCTGGCTCA	ATCAAGGACA	3540
TGTAAGCCGC	ACAGTTAGCT	TGGCTTCCTG	AATGTGGTTG	AACATTGGCA	AATTTAGCAC	3600
CGAAAATTC	TTTTGCGCGT	TCAATAGCAA	GAGTCTCTAC	AACGTCTACT	ACATCAGTTC	3660
CACCATAATA	ACGGCGTCCT	GGGTAACCTT	CGGCATATTT	ATTTGTCAAG	ATAGACCCTT	3720
GAGCTGCCAT	AACAGCCTTG	GAAACTACGT	TTTCCGAAGC	AATTAACCTG	ATATTATTTT	3780
GTTGGCGTTC	TTCTTCTTTG	GCAATAGCAT	TCCAGAGATC	AGCATCATAT	GCTTTAAAAT	3840
CATCTTTGTC	AAAAATCATA	GGTCTTCTCC	TTTATTGTGT	GACTAGTCCA	TTAGTTTGAT	3900
TTTACAATAA	GAAAATCAAA	CTAACAGATG	CGAATAAACC	GTTTCTGCAT	TTTATCACAA	3960
GTATAGCCAA	CTTTTTCATA	AAATGCATGA	GCACCCAGAC	GATGATTGGC	AGAATTTAAG	4020
CGGATAAACC	CATAACCACA	TCTTTTTCCT	TCTTCTTCCA	ACCCTTGTAG	TAAACTTTTA	4080
CCAATACCTT	GACCTTGCGC	TTGAGGTGAA	ACTGCTAAAG	CTAAGATATT	AAATCCTGCT	4140
TTGGAATAGA	GTGATTGCTA	AACCTTCAGC	TGGACATATC	CAAGTAAGAC	ATGATTAGCT	4200
GCATCCTCAT	AGCCAAGTAG	GAAATGATGG	GAATCCTGAG	ACAGTCTAGC	TAGTTGGCTA	4260
GCCGTTTCCT	CTGGACTAAA	AGTATAACCC	AAAGCCTCTT	GTTTGATCTC	ACATATAGCT	4320

711

TTCACATCAG TTTCTCTTAA ATCTCTTAGC ATCTCATTCC TCCTCAAAAG AAATCTTTGG	4380
CAACCGAGCA AGAATATCTT CTCGCTTAAT GGCCCTTGA CGTAAGATTT TCACCTTGTC	4440
TCCCACAAA TTCAAAATAG TTGAATCCTG TCCAGTTAGA AAAGCATCGT CTTCCAGACC	4500
CAGAACCTCT TGGTCAAAAT CCTCTAGAAT TTGATTAAAG GTCACCTCCAC TCGCCTGACC	4560
TGAGATATTG GCAGACGGCC CAATCAAGGG ACCTGTCTCT CGAATCAAAT CAAGGGTAAT	4620
GGGATGACTA GGCATCCGAA ATCCAACAGT TGCAAGGCCA GAATTGACCC AATAGGGAAC	4680
TCCGTCAATTA GCTTCGAGAA TAATGGTCAA GGGACCTGGT AAAAAGATCT CTACAAGTTT	4740
TTGAAGATAA GTTGGCTGAT TCTTTGAAAA GTACAAGATG TCCTCTAAAG AGGCAACATT	4800
GAGATTGAGC GCCTTGCTCT TACGTCGACG TTTAAGCTGG TAAACATGGT CAACTGCTTT	4860
TTCGTCTAGC GCCTTAGCAA AGAGACCGTA AACTGTCTCT GTAGGCAAAA CGACAGCTCC	4920
ACCATTTTCC AACTCTTGTC TAATCCTGTC CATCATCAAC GACAACCATC CTATCTTGAC	4980
CAAATTGGTC CTTGAGTGTT CGTACTCGCT TTTCAGGAAG ATGTTTCCTA AAAAGTTCAG	5040
GAACACTTIG ACCTTGCTTG TATCCAATTT CAAGGTAAAT CTTACCACCA TCTTTGAGAT	5100
AGTCTTTTGC ATCTTCCGCA ATTCTACGGT AAATAGCTAG GCCATCCTCA TCTGCAAAGA	5160
GAGCTAGATG AGGCTCCGAA TACAAGACAT TCAAGCCTAC CTCTGACTCA TCTTCACGAG	5220
AGATATAGGG TGGATTGGAA ACAATTATAT CATATTTTTC AGAAATTTCT GTAAAACACT	5280
CAGATTTTTT TAAAAATATT TGAAGATTTT GATTTTTAGC ATTTTCGCTA GCTACATCTA	5340
AAGCATCTTG GGAAATATCT GCTGCCGTCA CTGACCAATC TGGTCTGTTT TTTGCTAGAG	5400
CGAGAGCAAT AGCTCCACTA CCTGTTCCGA TATCTAGGAC CATAAGATTT TTCACAGGAT	5460
TTTCAGCCAG GATAAGCTCC ACCAACTCCT CTGTTTCTGG ACGAGGAATC AAAACCCGTT	5520
CATCCACCTT TAAATGCATT CCATAAAAAT CTGCCTGTCC AATGATGTAC TGAGCTGGCT	5580
TGTGAGCTGC TAGTTGCTGG TAAATATCTT CTACAAATTG TTTTCTTCC TCTGTTGTCA	5640
CCTCCTGCTG GAGGGCAAAA ATAAAGTCTG TAAAGATAG ATTTTTCAGA CTACGATAGA	5700
CAAAAGAGAG GCTTTCGGCT TCCTCTCCTT GTCTTATCAA CTCTTCTCA AAATTTGAAA	5760
ATAATTGAGC TAATTTGATT ATTTGTTTAA TTCTTCTAGT TTTTGTGTTT GGTCAATAAG	5820
CACCAAGGCA TCCACAACCT CGTCCAATTT ACCAGACAAA ATCGTATCTA GTTTTGGAG	5880
GGTCAAGCCG ATACGGTGGT CTGTGACACG GTTTTGTGGG AAGTTATAAG TTCGGATCCG	5940
TTCTGAACGG TCACCAGTAC CGATTGTGCA CTTACGCTCA GCGTCTGCT CATCTTGAGC	6000
AATCTGAGCA AAGTGGTCAG CAACACGGGC ACGGATGATT TTCATGGCCT TCTCACGGTT	6060

712

CTTCTGCTGG	GTACGTTCTT	CCTGCATCTC	AACCTTGATA	TTGGTTGGCA	AGTGAACGAT	6120
ACGAACGGCA	GTCGCAACCT	TATTGACGTT	CTGTCCACCA	GCACCAGAGG	CGTGATAGAT	6180
GTCGACACGA	AGGTCTTTTG	GATCAATGTC	GTATTCAACC	TCTTCAACTT	CTGGCATAAC	6240
AAGAACTGTC	GCTGTGGAAG	TATGAACACG	GCCTTGGCTT	TCTGTCACAG	GAACACGTTG	6300
CACACGGTGG	GCACCTGATT	CATACTTAAG	CTTAGAGTAT	ACAGACTGAC	CTGAAACCAT	6360
AGCAACCACT	TCTTTAAAAC	CACCGACACC	ATTCATAGAG	GCTTCCATGA	CTTCAAAGCG	6420
CCAACCTTGG	GCTTCCGCAT	ACTTTTGGTA	CATAGTTAGC	AAATCTCCAG	CGAAAAGTGC	6480
CGCTTCGTCT	CCACCAGCTG	CTCCACGGAT	TTCAAGGATG	ATATTCTTGT	CATCGTTTGG	6540
ATCCTTTGGA	AGGAGCAAAA	TTTTCAGTTT	TTCTTCATAT	TCTTCTTTTT	CAGCCTTGCC	6600
ATCTTTGAGT	TCTTGCTTGG	CCAATTCTTC	CAAGTCCGCA	TCTCCGCCTG	ATTCCTTAAT	6660
CATCTCTTCG	GCATCGACGA	TATTTTGAAG	GACTTGTTTA	TACTCACGGT	AGGCTATTAC	6720
GGTGTACGA	TTGGAAGCTT	CTTCTTTTGA	AAGCTCCATA	AAACGCTTGG	TGTCTGAAAC	6780
GACATCAGGG	TCATCAGCA	ATTCTCCTAA	TTCTTCATAA	CGGTCTTCTA	CAACTTGTAG	6840
TTGATCATAG	ATGTTCAATT	TTTCTCCTTA	TTTCTCAATT	GTAAATCAT	AGATTGCTAC	6900
TACTTCATTG	TCGGATATTT	CCCCAGTTTC	TTTAAATCCA	TAACTGAGGT	AACAAAATCT	6960
TGCTGTGTTCA	TTTTCTGGTT	CATAGGACAA	CCAAAGTTTA	TTGCTTAAAC	CTGCTGGCGC	7020
TGTTGGAACA	TAGTCTAGTA	CTTTATCCAT	AATTGGTTTA	AAATATCCTT	GATTTTGAAA	7080
ATTCTTATCA	ATCATAAAAC	GAAATAGTAA	ATAATTTCCA	CTACTAATTC	CGATCTTTTT	7140
ATCATAAGCT	ATCATCACAA	AACCTATAAT	TGCATCATTA	TCATAAACTG	CCAATGGAGC	7200
TACAAAATCT	CCATTTTATG	TGTAGACGTA	TGCTTCAGCT	AAACTAATTG	CGTTGGTTGC	7260
AATGAATTGT	TTTTGATATT	CCTTGACATC	CAAATTTAAA	ACATCAAAAT	AATTTTCCAT	7320
TGTAACATCT	CTTAGTTCAA	TTGTCATAGT	TTTGCTCCTT	GTTAGAGGTT	ATCATTGGCG	7380
CAAAATAATG	TTTACGGCAA	ACTGAGATAT	AGGTTTCGTT	ACCACCAATC	TGGATCTGTT	7440
CTCCATCGTA	AACGGGCAGT	CCATCCTGTG	TTGCAACAC	CATGGTCGCC	TTTTTCTTGC	7500
AATACTGACA	GATGGTCTTG	ATTCGTCAA	TCTTGTCTGC	TAAAAGCAAG	AGATATTTGG	7560
AACCTTCGAA	CAATTCATTG	CGAAAGTCAT	TTTTCAAGCC	AAAAGCCATG	ACGGGTATGT	7620
CTAACTCGTC	CACAACACGA	GCTAGGTCGT	AAACATGGTG	GCGTTTGAGA	AACTGGGCTT	7680
CATCGACCAA	AACACAGTAA	GGTTTTTCTG	GTAGGTCCTG	GATATAGCCA	AAGATATCCG	7740
TTGTTTCCTC	AATCGCAAGG	GCAGGGCGTT	TCATGCCAAT	TCGACTCGAC	ACATAGCCAA	7800
CGCCGTCACG	CGTATCCAGA	GCCGAGGTCA	TAATCACAAAC	ACCTTTTCCT	TGCTCCTCGT	7860

711

AGTTATAGGC CACTTTGAGA ATCTCAATCG TTTTACCAGA GTTCATGGTC CCATAACGAT	7920
AGTACAACTG TGCCATGTTT CTGCTTCAC GTCCATTCTT AAATTTTGC TACATTCTAG	7980
TATATCATAA TTTTCTTAAG CTMTAAACGG CAAAATGTGG TAAAATAGAA GAAATCAAAA	8040
ACTAGTGGAG GAAGCTATTA TGCCATTGTG ACGCATCGAT TTATTTGAAG GACGCACGCT	8100
CGAGCAAAAG AAAGCTCTTG CTAAGGAAGT AACGGAAGCA GTTGTCGCCA AACTGGAGC	8160
CCCTCAATCT GCTGTCCATG TCATCATCAA CGACATGCCA GAAGGAACTT ACTTCCCACA	8220
AGGGGAAATG CGTACTAAAT AAGCTAGCTT AAGCAGAATT GCTTAGGCTT TTTCAATCTC	8280
CAAGTAGCAT TCATTGAAGA AATATCCTAA ATTTGTTACA ATTTGAAAAG AAAGTGGAG	8340
AATTTCCAAG AAAAGAGCTA TTAATTAAG GAAACATTAT GATTACACGT GAATTTGATA	8400
CCATCGCTGC TATCTCTACT CCACTAGGTG AAGGGGCTAT TGGTATTGTC CGCCTGAGCG	8460
GAACAGACAG TTTTGCTATT GCGCAAAAGA TTTTAAAGG AAAAGACTTG AACAAGGTTG	8520
CCAGCCACAC TCTCAACTAC GGTCACATTA TTGATCCTCT GACTGGTAAA GTCATGGACG	8580
AGGTTATGGT TGGGGCTATG AAGTCTCCAA AGACCTTCAC TCGTGAGGAT ATTATCGAGA	8640
TTAACACCCA CGGTGGGATT GCGGTGACCA ATGAAATTCT CCAGCTAGCT ATTCGTGAAG	8700
GGGCTCGGTT GGCAGAACCT GGTGAATTTA CCAAACGTGC TTTTAAAC GGTGCGTAG	8760
ACTTGACACA GGCAGAGGCT GTGATGGATA TCATCCGTGC CAAGACTGAC AAGCCCATGA	8820
ACATTGCGGT CAAACAATTA GACGGCTCCC TTTCTGACCT CATTAAACAT ACCCGTCAAG	8880
AAATCCTCAA TACACTTGCC CAAGTTGAGG TCAATATCGA CTATCCTGAG TATGACGATG	8940
TTGAGGAAGC CACTACTGCT GTTGTCCGAG AGAAGACAAT GGAGTTTGAG CAATTACTAA	9000
CCAACTCCT TAGGACAGCA CGTCGTGGTA AAATCCTTCG TGAAGGAAT TCAACGGCTA	9060
TCATTGGACG TCCCAACGTT GGGAAATCAA GCCTTCTCAA CAACCTCTTG CGTGAGGACA	9120
AGGCTATCGT AACAGATATC GCTGGGACAA CACGAGATGT CATCGAAGAG TACGTCAACA	9180
TCAATGGTGT ACCTCTCAA TTGATTGATA CAGCCGTAT TCGTGAAACG GATGATATCG	9240
TTGAACAAAT TGGAGTTGAG CGTTCGAAAA AAGCTCTTAA GGAAGCTGAC CTAGTTCTGC	9300
TAGTACTAAA CGTAGTGAA CCACTAACCG CCAAGATCG CCAACTCCTA GAAATCAGTC	9360
AGGAGACTAA TCGCATTATT CTCTTAACA AAAGTACCT GCCTGAAACG ATTGAACTT	9420
CGGAACTACC TGAAGATGTC ATCCGCATTT CAGTTCTTAA AAATCAAAAC ATCGATAAAA	9480
TCGAAGAGAG AATCAACAAC CTCTTCTTTG AAAATGCTGG TTTGGTTGAG CAAGATGCTA	9540
CCTACTTGTC AAACGCCCGT CACATTTCTT TGATTGAGAA GGCCGTTGAA AGCCTACAAG	9600

714

CTGTTAACCA AGGTCTTGAA CTAGGGATGC CAGTTGACTT GCTTCAAGTT GACTTGACCC	9660
GTACTTGGGA AATTCTAGGA GAAATCACTG GAGATGCTGC TCCAGATGAA CTCATCACCC	9720
AACTCTTTAG CCAATTCTGT TTAGGAAAAT AAGAAAAATC CATGATCCTT CATTTCGGTCA	9780
TGGATTTTAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT TATGGAGCCT	9840
ATTTTATTGT AGAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA AACAACTCAT	9900
TAGAAAGAAT CATATGGAAC AATTACATTT TATCACAAAA TTAGTAGACA TTAAAGACCC	9960
TAATATCCAG ATTTTAGACA TCATCAATAA GGATACACAC AAGGAAATCA TCGCCAAACT	10020
GGACTACGAC GCCCCATCTT GCCCTGAGTG CGGAAACCAA TTGAAGAAAT ATGACTTTCA	10080
AAAAACCTTC TAAAATTCCT TATCTTGAAA CGACTGGTAT GCCCACTAGA ATTCTCCTTA	10140
GAAAGCGTCG ATTCAAGTGC TATCACTGTT CAAAAATGAT GGTGCTGAA ACTTCTATCG	10200
TCAAGAAGAA TCACCAAATC CCTCGTATCA TCAACCAAAA GATTGCTCAA AAGTTAATTG	10260
AAAAGATTTT TATGACTGAT ATTGCCCATC AGCTTTCCAT CTCAACTTCA ACTGTTATTC	10320
GTAAGCTCAA TGACTTTCAC TTTAAACATG ATTTTCTTG TCTTCCTGAG ATTATGTCTT	10380
GGGATGAGTA TGCTTTTACA AAAGGGAAGA T	10411

(2) INFORMATION FOR SEQ ID NO: 90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2393 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GTTTTGGGTT CTGGAAATTA TCAGATGGTT GGAAAAGCCG TCCACATCAA GATAGTGTTT	60
GGAGATTTAA GTTTAAATTG AAGAACTAA CACAGAGGAA ATGGAGTATA GACCTAACAA	120
GACGTATTGA GCAACTGAAT TTGTCTATTC GAGGATGGAT AACTATTGC TCATTGGGAA	180
ATATGAAAAG TATAGTCGCC AGCATAGATG AGCGCTTGCG TACTCGCCTA CGAGTGATT	240
TCTGGAAGCA ATGGAAGAAG AAATCGAGAC GATTATGGGG ATTGCTTAAG TTAGGAGTTC	300
CTAAATGGAT AGCAGATAAG GTATCTGGCT GGGGCGACCA TTATCAATTA GTAGCTCAGA	360
AGTCGGTACT TAAACGTGCT ATATCAAAAC CAGTCCTGGA AAAACGTGGA CTGGTTTCGT	420
GTTCGATTA TTACCTTGAA CGACATGCGT TAAAGTTAG TTGAACCGCC GTATGCCAAA	480
CGGCACGTAC GGTGGTGTGA GAGGGGCTAG AGATTATCCC CTAATCGATT AACTCCCCTG	540
AAATTTATTT TAATTATGCA AATTTCACGT ATTTTGTATG CTGAGACGAC GATCCTGGGA	600



715

ACTTTTCAGA TATTTTTTTG ACTATCTAAA TCTATCATTA GAAAAGCTTA GAGCGCCAAA	660
GGATTTCAGC GTTTTTCTGA TTTTAAAGAC TTTTCCAGT CTCTTTTTCG ATTGAAGATG	720
TAATTATTCT ACTAACTAAC TAACCTCTTA GTACTAGCCA ACAACGATAA TCATAATTCC	780
TCCTAAAATT AGGAATAATA AAGGCAATAG TTTTGTGTTT TTCATGTAAA AAACCTCACT	840
TTTGTGTTCT GCTATTTTAT GCTAAAATAT TAAAAATCAA ATTTAATTCC AAAGTTTGTA	900
ACTAAAGGGG GAGCGCTACA TGTCTAATTC ATTTGTCAAG TTGTTAGTCT CTCAATTATT	960
TGCAAATTTA GCAGATATTT TCTTTAGAGT AACAATCATT GCTAACATAT ACATTATTTT	1020
AAAATCAGTA ATTGCCACAT CACTAGTTCC TATCTTAATA GGAATATCCT CTTTTGTTGC	1080
GAGTCTTTTA GTTCCGTTGG TTAATAAAG GTTAGCGCTA AATAGGGTTT TATCTTTATC	1140
TCAATTTGGA AAGACTATAT TATTGGCGAT ACTGGTAGGA ATGTTTACCG TAATGCAATC	1200
CGTAGCGCCT TTGGTGACCT ATCTATTTGT TGTGCAATT TCCATACTAG ATGGTTTTCG	1260
AGCACCCGTT TCCTATGCTA TTGTGCCACG CTATGCGACC GATTTGGGTA AGGCTAATTC	1320
AGCCTTATCA ATGACTGGTG AAGCTGTTCA ATTGATAGGT TGGGGATTAG GTGGACTCTT	1380
GTTTGCAACA ATTGGTCTGT TACCTACCAC GTGTATCAAT TTAGTCTTGT ATATCATTTT	1440
TAGCTTTCTG ATGTTATTTT TCCCTAACGC TGAAGTGGAG GTGTTAGAGT CAGAAACTAA	1500
TCTTGAAATT TTGCTCAAAG GTTGGAAGTT AGTTGCTAGA AATCCTAGAT TAAGACTTTT	1560
TGTATCAGCA AATTTATTGG AAATTTTTTC AAATACGATT TGGGTTTCTT CCATTATACT	1620
TGTTTTTGTA ACGGAGTTAT TAAATAAAAC GGAAAGTTAC TGGGGATATT CTAATACAGC	1680
ATACTCTATT GGTATTATAA TTAGTGGCTT AATTGCTTTT AGGCTATCTG AAAAGTTCCT	1740
TGCTGCTAAA TGGGAACCCC AATTATTCAC CCCAAATCTA AAAACCATCC AGAATCCTTG	1800
CCTTAGCTTA GATCCTGGAT GGTTCCTTTT TTCACCCAAT GGGTGTGTTT TACTAGACAA	1860
AAAAGAGTTT CCCCTTTATG GTATAAGTGT AGAAAAAAC ACAAAAAGAA AGGAAACTCA	1920
CATGAACAGT TTACCAAATC ATCACTTCCA AAACAAGTCT TTTTACCAAC TATCTTTCGA	1980
TGGAGGTCAT TTAACCCAGT ATGGTGGTCT TATCTTTTTT CAGGAACTTT TTTCCAGTT	2040
GAAACTAAAA GAGCGGATTT CTAAGTATTT AGTAACGAAT GACCAACGCC GCTACTGTCTG	2100
TTATTCGGAT TCAGATATCC TTGTCCAGTT CCTCTTTCAA CTGTAAACAG GTTATGGAAC	2160
GGACTATGCT TGTAAGAAT TGTACGCTGA TGCCTACTTT CCAAAATTGT TGAAGGAGG	2220
GCAGCTTGCT TCACAGCCAA CCTTATCCCG TTTTCTTTCC AGAACTGACG AGGAAACAGT	2280
CCATAGTTTG CGATGCCTCA ACCTTGAATT GGTCAATTC TTTTACAGT TTCACCAGCT	2340

716

AAACCAACTC ATTGTAGATA ACGATTCTAC CCATTTCCACA ACTTATGGCA AGC

2393

(2) INFORMATION FOR SEQ ID NO: 91:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4762 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TTTGATCTT TTTAGGTCTC TTTCAATCCA AACCCCTTAA ACTATACGTC ATTTTCGGTTC	60
CTGCAAGTCT TGTGGTAATT TTAGGTTTGA TTTTACTTTT CTTTTCACAA GAGCCTCTGC	120
ACGCTTCTTA TTTGATGGTC GTCTTCCCTG TTTTCTACT TTTATTGGTA ACCAATATTA	180
AGAGTCAACA GAGGGGGCGT AGTGCTAGAA GAAGCCGAAG AGAAACGCCA TTATGCCTAT	240
GGAGTCGTTT CTTCAAAGGA AATCTATATC TGCTAGTTTT TGGGTTTGTC TATCTTTTGT	300
CTGTTCTTTT TTTGATGAAG TTTGTCCTTT ATCCAGTACC TTATCAAGAA CGTAATCGTC	360
TTGCTGATTT GGTAAAAGAG GAGACAAATA CGGAAGATGC TATCTCATGC ATGGGATGAT	420
ACTGCGACTC TTTATCGTAA GAGTGAGCGC TTGTCCCATC GGCGATTTTG TCCCCGTTGC	480
ACTATACAGC AACTGAGGAA AATCGTAATA AGTTACTTAA TGACTTGAAA GAAAAACAAC	540
CTAAGGTGAT TGTGGTAAAT GATAAGGTGG TAGTCTGGTC TGAAGTGGAA ACACTCTTAA	600
AAGAAAATTA CCAACAAGTA AAGACTGATT ACTCAGAGTT TAAAGTCTAT AAAATTAAAT	660
AACCAAATCA ATATCTTGTG TATTTTTTAA AATTTTAGGA TTTTAAACAC AAGATATTGA	720
TTTTTCTTTT TAGAGTGGTA TAATACTTTT TAGAAAGAAC ATTTTAGAAA AGAGCATGCA	780
TATGATTGCA CTAGAAGAAA AAATTACAAT TTTGCCAACT CTCTTCGTCG AGAAACGAGA	840
TGGGAGACGT GTTGATTTG ATGTGGACAA GATTGACAAG GCTCTCCACA AGCGGGCTGA	900
CAAGGTTATG GATGTGACAC CCCTGGTTGA AAAATGCCTC AATGATCTGA CTGAGCGAAT	960
TATTACAGAA ATTCATAGTC GCTTTCCACA GGGAATTAAG ATTTACGAAA TTCAAAATAT	1020
CGTAGAACAT GAACTCCTTG AAGCCAAAGA ATATGCGCTG GCTCAGGAGT ATATTACTTA	1080
TCCGACACAG AGGGATTTTG AGCGCTCAAA AGCGACGGAT ATCAACTTTA GTATTCATAA	1140
ACTTCTCAAC AAAGACCAGA CAGTTGTCAA TGAAAACGCT AATAAAGACA GTGATGTCTT	1200
TAACACTCAG CGTGATTTGA CAGCAGGGAT TGTGGGAAA TCAATCGGAC TGCAATGCT	1260
TCCTAAGCAC GTAGCCAATG CCCACCAAAA GGGGGATATC CACTATCAGG ATTTGGACTA	1320
CAGTCCCTAT ACCCCTATGA CCAACTGCTG TTTGATTGAT TTAAAGGGTA TGTGGAAAA	1380

717

TGGTTTTAAG ATTGGAAATG CAGAGGTAGA GAGTCCCAAG TCTATCCAGA CTGCGACAGC	1440
ACAGATTCT CAAATCATTG CCAACGTTGC TTCTAGCCAG TACGGTGGCT GTTCAGCTGA	1500
CCGTATCGAT GAAATTTTGG CGCCTTATGC ATAGAAGAAT TATCAAAAAC ATCTCAAAGA	1560
TGCAGAAGAG TGGGTATTGC CTGAAAAACA GGAAGATTAC GCTTGAAGA AAGCGCAAAA	1620
GGACATCTAC GATGCCATGC AATCTCTTGA GTATGAAATC AATACTCTCT TCACTTCAAA	1680
TGGACAAACA CCTTTTACTT CGTTAGGTTT TGGTCTGGGA ACCAGTCGTT TTGAACGAGA	1740
AATTCAAAAA GCTATTTTAA ACATTCCGAT CAAGGGTCTT GGTTCAGAAC ACCGTACGGC	1800
TATCTTTCCT AAACCTATCT TTACGCTTAA AAGAGGCCTC AACTTAGAGG AAGGAACTCC	1860
CAACTATGAC ATCAAGCAGT TGGCTCTAGA GTGTGCAACC AAGCGGATGT ATCCAGACGT	1920
CTTGCTTAT GATAAGATTG TTGATTTGAC AGGTTCTTTC AAGGTGCCTA TGGGCTGCCG	1980
TTCTTTCCTT CAAGGGTGGA AGGATGAAAA TGGTGTAGAA GTCAATTCAG GTCGCATGAA	2040
TCTGGGTGTT GTGACGGTTA ATCTGCCTCG TATTGCTCTT GAGTCTGAAG GTGATATGAA	2100
TAAGTCTCG GAAATCTTCA ACGAGCGAAT GAATATCGCA GAAGATGCTC TTGTTTACCG	2160
TGTCGAACGC ACTAAAGAGG CGACACCAGC GAATGCTCCT ATTCTTTATC AGTACGGTGC	2220
TTTTGGCCAT CGTCTAGGTA AAGAAGAAAG TGTGACCAG CTCTTTAAGA ATCGTCGTGC	2280
GACCGTTTCG CTGGGCTATA TCGGCTTGTA TGAAGTAGCG ACAGTTTCTT TTGGTAACAG	2340
CTGGGAAAGT AATCCAGATG CTAAGGAATT CACGCTAGAC ATCATTACG ATATGAAACG	2400
CCGTGTAGAA GAGTGGTCAG ACCAATATGG CTACCATTTC TCTATCTACT CAACACCATC	2460
CGAAAGTCTG ACAGACCGTT TCTGCCGACT AGATATAGAC AAGTTTGGCT CTATTCTCTA	2520
TATCACAGAC AAGGAATACT ACACCAACTC TTTCCACTAC GATGTTTCGTA AAAATCCAAC	2580
ACCGTTTGAA AAATTGGACT TTGAGAAAGT CTATCCGGAA GCAGGTGCGT CAGGTGGTTT	2640
CATCCATTAT TGTGAGTATC CAGTCCTTCA GCAAAATCCA AAGGCCTTGG AAGCTGTCTG	2700
GGATTATGCT TATGACCGTG TAGGCTATCT AGGCACCAAT ACTCCGATTG ACCGTGCTA	2760
CAAGTGTGAC TTTGAAGGGG ATTTTGAACC AACTGAGAGA GGGTTTGCTT GTCCAAACTG	2820
TGGCAATAGC GACCCTAAAA CAGTAGATGT GGTGAAACGA ACTTGTGGCT ACCTAGGTAA	2880
TCCTCAAGCA AGACCGATGG TCAACGGGCG TCACAAGGAA ATCGCTGCGC GTGTCAAACA	2940
TATGAATGGT TCAACGATTA AAATAGCTGG GCATCAAGTA ACAAATTAGA AAGAAATGAA	3000
ATGGGAAAAT ATCAACTAGA CGATAAGGGG CGCGCACAAG TGACCCGTTA TCACGAGAAA	3060
CACTCTAAAG GTGGAGCTGG TAAGAAAGAA CGCTTGCTTA GCTTCAGAGA ACAATTTTAA	3120

718

AACAAGAACA AGAAAAATA AAAGTGAGAG CCAGCTCTCG CTTTCTCAT AGTGGGAGGT	3180
AAGGATGGAA TTACGCAGAC CAAGATTAGC GGATAAGAAA GCTGTTTATG ATATGATGAC	3240
AGAGTTTGAA AAATTTCAGT CGCCTCACGA CGGCGGTTTC TGGGATACAG AGAACTTTGT	3300
GTATGAAGAC TGGTTAGAAA GCAATCAGGA ACAGGAAATG GGGATTAAATC TGCCTGAAGG	3360
ATGGGTTTCT GCAATTTCAGT TAGTGGCTTT TTCTGAGAAA GGTCAAGCAG TTGGATTCTT	3420
TAATCTCCGG TTGCGCCTCA GTAACCTTCT ACTAGAAGAA GGTGGCCACA TTGGCTACTC	3480
CATTCGTCCA TCTGAAAGAG GCAAGGGTTA TGCAAAAGAG ACTCTCCGTC AGGGCTTGCA	3540
AGTTGCTAAG GAAAAGAACA TCAAGAAAGC TCTGGTGACC TGTAGTGTGA ATAATCCTGC	3600
TAGCAGAGCA GTCATTCTAG CAAATGGTGG AATATTTGAG GATGCTCGCA ATGGAGTCGA	3660
GCGTTATTGG ATAGAGGTAG CGAATGAATA ATCCAAAACC ACAAGAATGG AAAAGCGAGG	3720
AACTTAGTCA AGGTCGTATC ATTGACTACA AGGCCTTTAA CTTTGTGGAC GCGGAAGGCG	3780
TGCGCAACTC TCTCTATGTA TCAGGCTGTA TGTTCCTACTG CGAGGGATGT TATAATGTTG	3840
CGACTTGGTC TTTAATGCT GGCATTCCCT ATACAGCAGA ATTAGAAGAG CAGATTATGG	3900
CAGACCTTGC CCAACCCTAT GTTCAAGGCT TGACTTTGCT GGGAGGGGAG CCTTTTCTCA	3960
ATACTGGGAT TCTCTTGCCA CTTGTTAAGC GGATTCGGAA GGAATTGCCA GACAAGGACA	4020
TCTGGTCCTG GACCGGCTAC ACTTGGGAAG AAATGATGTT GGAAGCTCCA GATAAACTGG	4080
AATTCTTGTC ACTGATTGAC ATTCTGTGCG ATGGAAGATA TGATCGAACT AAGAGAAATC	4140
TTATGCTCCA GTTTCGAGGT TCATCTAACC AACGAATTAT CGATGTGCAA AAATCGCTCA	4200
AAAGTGGGCA AGTAGTGATT TGGGACAAGC TCAATGACGG AAAAGAAAGC TATGAACAGG	4260
TGAAGAGAGA ATGAAGAAAA AGGACTTAGT AGACCAACTA GTCTCAGAGA TCGAGACGGG	4320
GAAAGTCAGG AACTGGGAA TATACGGTCA TGGAGCTTCA GGTAAATCAA CCTTTGCACA	4380
GGAATTGTAC CAAGCTTTAG ATTCTACTAC AGTAAATTTG CTAGAGACAG ATCCTTATAT	4440
CACCTCAGGA CGCCATCTGG TAGTACCCAA GGACGCGCCG AATCAAAAAG TGACAGCCAG	4500
TCTGCCAGTG GCGCATGAAC TGGAGAGTTT GCAGAGAGAT ATCCTTGCTT GCAGGCGGGT	4560
ATGGATGTCT TGACAATTGA AGAACCTTGG AAGGCTAGTG AGGTCTTGTC TGGAGCCAAA	4620
CCAATTTTGA TTGTGGAAGG GATGTCTGTT GGCTTTCTAC CCAAGGAACT CTTTGAAAAA	4680
ACCATCTGTT TCTACACGGA TGAGGAGACC GAATTAAAGC GACGCCTTGC TAGAGATACG	4740
ACTGTGAGAA ATCGCGATGC GG	4762

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

629

CGTATCTTCT CGTTCTGAAA AAGTTGGTGA GTACATCTCT GAAGCAATGG AAAAAGTTGG	780
CAAAGACGGT GTCATCACCA TCGAAGAGTC ACGTGGTATG GAAACAGAGC TTGAAGTCGT	840
AGAAGGAATG CAGTTTGACC GTGGTTACCT TTCACAGTAC ATGGTGACAG ATAGCGAAAA	900
AATGGTGGCT GACCTTGAAA ATCCGTACAT TTTGATTACA GACAAGAAAA TTTCCAATAT	960
CCAAGAAATC TTGCCACTTT TGGAAAGCAT TCTCCAAAGC AATCGTCCAC TCTTGATTAT	1020
TCCGGATGAT GTGGATGGCG AGGCTCTTCC AACTCTTGTT TTGAACAAGA TTCGTGGAAC	1080
CTTCAACGTA GTAGCAGTCA AGGCACCTGG TTTTGGTGAC CGTCGCAAAG CCATGCTTGA	1140
AGATATCGCC ATCTTAACAG GCGGAACAGT TATCACAGAA GACCTTGGTC TTGAGTTGAA	1200
AGATGCGACA ATTGAAGCTC TTGGTCAAGC AGCGAGAGTG ACCGTGGACA AAGATAGCAC	1260
GGTTATTGTA GAAGGTGCAG GAAATCCTGA AGCGATTTCT CACCGTGTTG CGGTTATCAA	1320
GTCTCAAATC GAAACTACAA CTTCTGAATT TGACCGTGAA AAATTGCAAG AACGCTTGGC	1380
CAAATTGTCA GGTGGTGTAG CGGTTATTAA GGTGGAGCC GCAACTGAAA CTGAGTTGAA	1440
AGAAATGAAA CTCCGCATTG AAGATGCCCT CAACGCTACT CGTGACGCTC TTGAAGAAGG	1500
TATTGTTGCA GGTGGTGGAA CAGCTCTTGC CAATGTGATT CCAGCTGTTT CTACCTTGGT	1560
ATTGACAGGA GATGAAGCAA CAGGACGTAA TATTGTTCTC CGTGCTTTGG AAGAACCCGT	1620
TGCTCAAATT GCTCACAATG CAGGATTGTA AGGATCTATC GTTATCGATC GTTTGAAAAA	1680
TGCTGAGCTT GGTATAGGAT TTAACGCAGC AACTGGCGAG TGGGTTAACA TGATTGATCA	1740
AGGTATCATT GATCCAGTTA AAGTGAGTCG TTCAGCCCTA CAAAATGCAG CATCTGTAGC	1800
CAGCTTGATT TTGACAACAG AAGCAGTCGT AGCCAATAAA CCAGAACCAG TAGCCCCAGC	1860
TCCAGCAATG GATCCAAGCA TGATGGGCGG GATGATGTAA GCTTTCTATA GAAAACAACT	1920
TATAAAAAAC ACAAAAGGAG GGAATGACTA ACCCTTCTTT TTATAGGCTC TTTGCTCACT	1980
GTAGTGGGTT GAAGTCAGCT AAGCTCGAGA AAGGACAAAT TTCGTCCTTT CTTTTTTGAT	2040
GTTCAAAGCG ATAAAAATCC GTTTTTTGAA GTTTTCAAAG TTTTCGAAAC CAAAGGCATT	2100
GCGCTTGATA AGTTTGATGA GATTATTGGT CGCTTCCGGT TTGGCGTTAG AATAGTGTAG	2160
TTGAAGGGCG TTGATAATCT TTTCTTTATC TTTGAGGAAG GTTTTAAAGA CAGTCTGAAA	2220
AATAGGATGA ACTTGCTTAA GATTGTCCTC AATAAGTCCG AAAAAATTCT CCGGTTCTTT	2280
ATTCTGAAAG TGAAACAGCA AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTTGTA	2340
ATAGCTCAAA AGCTTGCTA AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGACG	2400
ATAAAATCGC TTATCACTCA GTTTACGGCT ATCCTGTTGT ATGAGCTTCC AGTAGCGCTT	2460

630

GATAGCCTTG TATTCATGGG ATTTTCGATC CAATTGGTTC ATAATTTGAA CACGCACACG	2520
ACTCATAGCA CGGCTAAGAT GTTGTACAAT GTGAAAGCGA TCCAACACGA TTTTAGCATT	2580
CGGGAGTGAA ACAGTCTGGG AGACTGTTTC AGCCTGAGCC TAGAAATTTG AAAGCGAAGC	2640
TGTTTAGCCA AGTCATAGTA AGGACTAAAC ATATCCATCG TAATGATTTT CACTTGACAA	2700
CGAACGGCTC TATCGTAGCG AAGAAAGTGA TTTCGGATGA CAGCTTGTGT TCTGCCTTCA	2760
AGAACAGTGA TAATATTAAG ATTATCAAAA TCTTGCGCAA TGAAACTCAT CTTTCCCTTA	2820
GTGAAGGCAT ACTCATCCCA AGACATAATC TTTGGAAGCC GAGAAAAATC ATGCTCAAAG	2880
TGAAAGTCAT TGAGCTTGCG AATGACAGTT GAAGTTGAAA TGGCCAGCTG ATGGGCAATA	2940
TCAGTCATAG AAATTTTTTC AATTAACTTT TGAGCAATCT TTTGGTTGAT GATACGAGGG	3000
ATTTGGTGAT TTTTCTTTAC CAGGGGAGTC TCAGCAACCA TCATTTTTGA ACAGTGATAG	3060
CACTTGAAAC GACGCTTTCT AAGGAGAATT CTAGAAGGCA TACCAGTCGT TTCAAGATAA	3120
GGAATTTTAG AAGGTTTTTG AAAGTCATAT TTCTTCAATT GGTTCGCCA CTCAGGGCAA	3180
GATGGGGCGT CGTAGTCCAG TTTGGCGATG ATTTCCCTGT GTGTATCCTT ATTGATGATG	3240
TCTAAATCT GGATATTAGG GTCTTTAATA TCGAGCAGTT TTGTGATAAA ATGTAATTGT	3300
TCCATATGAA TCTTTCTAAT GAGTTGTTTT GTCGCTTTTC ATTATAGGTC ATATGGGACT	3360
TTTTTTCTAC AACAAAATAG GCTCCATAAT ATCTATAAGG GATTTACCCA CTACAAATAT	3420
TATAGAGCCG AAAATTCACA TCTAATATAT GCAGACTACT TTGAAATGAA ATTAAAAAAA	3480
TTATTAAAGG ATGACACAAA AGTTTTTGAA AAATCTACAT TCAAATTTGT AGAAGGATAT	3540
AAAATATACC TGACAGAATC TAAAGAATCT GGAATTAAAC AAATGGACAA TGTCATAAAA	3600
TATTTTGAGT TTATTGAATC TAAAAGTATT GCTTTATATT TTCAAAAACG ATTAAATGAG	3660
CTGATAGATT AAATAGCATT TTCTCTGTTG AGATATTGTT TTTAAAATAT TGTACTAAAT	3720
GATTGATGCT ATGTGGAAAT ACAAAAAAAT GTTTTTGATA CGAAGTTGAC CTGTATTTTT	3780
TATACTAATC ATTTTCGTAT TTTTGTATT AAACGATATA AGTTTGTGTG AAACCTTACA	3840
GGAATAAAGA CATTAAAAAA TAACAGTATA TCTATTTGTT TTATATATTT TACGAATTCT	3900
GCATAAATCT CTTTCTAGTA ATGTGTTGTA ACTCTGCTAT AATAGATTTA TTCTTTTTTG	3960
TGTTTACACA ATTTATTTTA TACTACCAAA AAAGGTCAGG ATTTTGTTCG TGACCTTTGA	4020
CAACTTTACC GATTCTTTAG TTCTACATAG CGCTTGTAAC AAATGTTTAC ATAGGCTTCT	4080
GAGAAAGGAC CACGTCCATT GTTAATCCAA TCAACAAGAA TTTTGACATG TTCTTTTAAA	4140
ATATAGTCCA AGTCATCAGA ATAATTCATT TTGCGTTTGT GACGCTCGTA CTCTTCAACG	4200
TCCAAGAGAC GTTTTCCCC ATCTGTAAAA ATTTTAACAT CCAAATCGTA ATCAATATAC	4260

631

TTCAGTGCTT CTTTATCCAG ATAGTAGGGG CTAGCCATAT TGCAATAGTA AGAAGTTCCA	4320
TTATCACGAA TCATGGCAAT GATATTAAAC CAATATTTCT TGTGAAAGTA AACAATAGCC	4380
GGTTCTCGAG TGACCCAACG ACGACCATCA CTTTCGGTAA CAAGTGTATG ATCGTTGACA	4440
CCAATAATGG CGTTTTCTGT TGTTTTTAGT ACCATGGTGT CCCGCCAAGT TCGGTGGAGA	4500
CTCCCATCAT GCTTATAACT TTGAATTGTA ATAAAGTCGC CTTCTTTTGG AAGCTTCATA	4560
ACTAACCAAC TTTCTACAAT TTATAAGTTT ATCATTACT ATTGTACCAT AAAATTACCC	4620
AAAATCTGTG AATTTCACCTT GGAAATATTA AAGATATTCT CTAAGAGCGC TTGCTATATC	4680
CGAAAAATCG TAGCCCTTTC GTGCTAAAAC TTGAGTTAAA CGCTGCTTCA GTTCGTATCC	4740
TTCATACTTT CGGGCATACT TAGTATATTG CTTATCAAGT TCCTTGAAGA TGAGTTCCTG	4800
AGTCGTTTTCT TCATCAACTT GACTATCCAA TTCGTCAAAG GCAATTTTAG CATCAAAATA	4860
AGAGAAGCCC TTGTTAGTCA AGTTCTGGAT AATCTTATCT TGCAGGGCAC GAGCTGGAAG	4920
TTTTCCCTCA TATTTTTTCA ATAGTTTATT GGCTACACGT TGAGCAACTT CCGAAAAATC	4980
AAAATCATTG AAGATTCTT CTATAGTAGA TTTTGAAATT CCTTTTTGTG CTAATTTCTG	5040
AGTCAGTACA TAAGGTCCCT TGTCTCCTGA AAGTTGATTG GCATTGATGA TAGCATAAGC	5100
GTACTGGCTA TCATTAAATCC ACTTCTCTTC TTAAAGATTA GCAATGACTT GAGAAACCAT	5160
GTTCATTA ATATCATATT TTTTCAGATA TTCTCTGACC TCTTTTTCAG TACGTGCTTT	5220
AAAGGATAAG TGGTAGAGGG CCAGATTCTT ACCATAAGAA AATTGAGCAA AGTCTTGAAT	5280
CTCTTTCAAT TCCTCTTCGC TTATCACCTT ATCTCTCGAT AACATAAAAC GAACAATTGT	5340
GTCTTCGGTG ATATAGCATT TGTCG	5365

(2) INFORMATION FOR SEQ ID NO: 78:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3636 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

TTTCCAGAAA GAAGTTGAGT AAAGTCTTTA TCAAAGAGAA TGAATCCGT ATTGGAAGTG	60
ACATTAGGTT TTATTTCTAC TTTACTAGCG TCCGCCCTAG CATTCTCTAA ATCTTTAATC	120
TCTTCTGTTG CCTATTTAT AGCCAGCTGA ATAACTGCTT GAGGATTTTC ACTCAGTCCA	180
TGAAGCTTAT CGTCCACCGA AGTATAAAGA CTCGAATGCA TCACTTGTA AATAATCAGA	240

632

GTCATTGTAG	AAAAAATCAG	GGTGAAGACA	CCGAAGTTGC	GGATAAAATA	ACTAAAGTCA	100
TCCGCATACC	ATGTTTTTTT	AAGTTTACTG	AACATCTTTT	AAAAGATACC	CAACACTACG	160
CAAAGTTTGC	AAATTCTCTG	CAAAAGTGGT	TCCCTTTAAT	TTCTTACGGA	CTTTTGAAAC	420
ATAGACTTCG	ACAACCGAAA	TCGTTGTATC	ACTATCAAAT	CCCCATAGAC	GGTCAAAAAT	480
CTGCGTCTTA	GGCAAAATCA	CATTTTGATT	TTGAAGGAAA	TAAACTAGTA	AATCGAACTC	540
TTTCCCCAGC	AATTCGACAG	GAGTATCTTC	AACTTTAACG	GTATTGGTTG	ATAAATTAAC	600
CACGATATTC	CCATAAGTCA	AGGTGTTTTT	ATTAACTTTC	CCTGAACGTT	TGAGAAGGGC	660
CTGAATCCGC	ATTTTAAGTT	CTTCTAGGTA	GAAAGGTTTG	GTCAGATAAT	CATCCGCTCC	720
CAGTTCAAAT	CCATGTCCCT	TGTCATCCAA	ACTTTCCTTG	GCAGTCATAA	TCAGAACTCG	780
TGTCGTAATT	CCCTTTTCAC	GCAATTCTTT	TAAGACTTGG	AAACCATTTT	TTTCTGGCAA	840
CATCAAATCC	AGCAAAATCA	AGTCATAGAC	ACCACTCTCA	GCTTCGTAGA	GACCTTCTTC	900
TCCATCAAAT	ACCTGCATAA	CATCCGCAAA	ATCGTCTAAA	AAGTCAAATA	CTGAATTGCA	960
CAGACCTAGG	TCATCCTCAA	CCAATAAGAT	TTTTATCATG	AGAAACTCCT	CCTTATTAAA	1020
ACTATTATAC	CAAATTTGCC	TTAAAAAATA	CTCAACTCTC	TGCATTTTAC	ATGAGATAGC	1080
TGAGTTTTCT	TTTTATTTTA	GGCTTATTTA	TGCATTTCCG	TATTGAAGAA	CAACTGCTTC	1140
GACTGCAGCT	TTTTACGGC	TAATCAAGTC	AACACGCGCT	GCAATTTCCCT	TGATTCCCAT	1200
ACCGATGTTA	CGGCTAAGAG	CAAGGTCAGA	AAGTTGCGGT	TCAAAGAACT	CCTTGTATTC	1260
CGCCAAGCGT	TGCTGAGTCT	TAAATACATG	AGCAGGAAGG	ATAACAAAGC	TATCAAAGCT	1320
CATATCTCCT	CCAAGGGCTG	CCTTAATCCA	AGCCCAGTTT	TCACGCGCCC	AAGACCAAGC	1380
TGTTTTCTGA	GTTGCTTGAT	GAGCTAGGAA	TTGGTAATAC	CAAGCAGACA	AGTCCTGTGG	1440
TTTGACCACA	AATTTGTCCT	TCCAAGAAGT	AATCAGGTTT	TGGATATTAT	CCGCATCTGT	1500
ACTGTATGCA	AGAGCTGCTG	CCAAGTGGCG	TTTAAAGACA	GCATCTGTTG	CGTGAGTATA	1560
AGTATCAAGA	TAAAGTGCTA	ACAAGTCTTT	AGTCTCATGA	TGTTTCATCT	CATTAATCAG	1620
AACTTGTGAG	CGAATAGCTG	CTGGGAGTCC	TGCAAGATTC	TCCTTGTGTG	TTGCGAAGAT	1680
TTGGCTAGCG	ACTTGACTAG	CTTCTGCATC	ATTTGAGCGA	ATCATCATCG	AAACAGCCAG	1740
CTGACGAACC	AATTCATCCT	CATCTGATTC	TCCGTCTTTA	GCTTCAAAAC	CAAGACGGTC	1800
ATAGTTATGA	CGAGCCAATT	TAGCAACCAG	TCCTTTGAAG	GCTGTTTCAG	CATCCGTTCC	1860
TTTCATCAATA	AAGCGCTCAA	GGGCTGAAAT	CACCTGAGAA	ACAGCTGAAA	CCACCAGATA	1920
AGACTCTTCC	TTAGCAAGTT	TATCAAGAAC	TGGAAGCAAG	TCTGCATAAG	AAATGTGCCC	1980
TGCCTCAGCC	AACAAACGAC	GTTCTTGAAC	AATTTGCACT	TTGCTTGTGT	TATCAAGTGT	2040



633

CTCTAGCTCA GCAAGAACAG CTGCTAACAA GTCTCCTTGA TAGTCGGTAA TATAGTGGGC	2100
AGTATTTTCA GTGTTGAGAC GAAGAGCTCC TTCATTTTCA GCAAGAAGAG CTGCGTAGCC	2160
AGGGATTTTCG ATACTTTTCAG TTTCGAGTGT ATCAGGCAAG CCTTTCCAGT TGCTATTGAG	2220
GGGCACCACC CAGAGACGGT TCTTGCTCTC GTTCTCACCG ATGAAGAATT GTTTTTGTGA	2280
AATCTTCAAG ACATCATTTT CAACTTTAAC AGTAAGAACT GGGTAACCAG GCTGTTCCAA	2340
CCAAGAATCC ATGAAGGCTG CGACATCACG TCCTGACGCT TGACCAAGGG CATCCCAAAG	2400
GTCACTACCA ATGGTGTTGC TGTATTGGTG TTTTTCAAAG TAGGCGTGCA AACCTTTAGC	2460
AAAATCAGCA TCTCCTAGCC AACGGCGAAG CATGTGCATG AGACGGCTTC CTTTGGCATA	2520
GACGATAGCG CCGTCAAAGA GTGTATTGAT TTCATCTGGA TGTTTAACTT CGACGTGGAC	2580
AGACTGAACG CCATCAGTAG CGTCACGTTT AAGAGCAAGA GGTACTCCAC CTGTTTGGAA	2640
ATCTTCAAAG ATATTCCAGC TTGGTTCGAT GGTATCCACA CAGACGTATT CCATCATATT	2700
AGCGAAACTT TCATTGAGCC AAAGGTCATC CCACCATTTC ATAGTCACGA GGTTCCTCAA	2760
CCATTGGTGA GCCAATTCAT GGGCCACAAC AAGGGCAACT TGTGACGGC TAGCAAATGT	2820
AGAGTTCTCA TCGACAACCA AGTAACTTC ACGGTAGGTC ACAAGACCCC AGTTTTCCAT	2880
AGCACCAGCT GAGAAGTCAG GAAGGGCGAT GTGGAGAGAT TGAGGAATTG GGTACTTAAC	2940
TCCATAGTAA TCTTCGTAAG ACTCGATAGA GCGAACAGCG ATATCCAGTG AGAAATCAAG	3000
ATTTGAAAGT GCATGTGCTT TGTTTGAGTA GACACCTACC AGGTACCAT TTTTAGTTTT	3060
AGCGGTCAAC CCTTGCAAAT CACCAGCAAC AAAGGCCAAC AAGTAAGAAG ACATGCGAGG	3120
TGTGTCTCA AACTTCCAGA TACCTGTTTC CTTACGGTTT TCAACATCGA TTTCTGGCAT	3180
GTTTGACAAG GCCAATTCAC CTTCTGCTTG GTCAAAGCGA AGAGAGAGGT CAAAAGTTGC	3240
TTTGGCTTCA GGCTCATCCA CACATGGGAA AGCTTCGGCG GCAAAATGGC TCTCGAACTG	3300
AGTAGACAAG ACCTCCTTCT TGACTCCATC AACTGTATAA TAAGAAGGGT AAATCCCTGT	3360
CATGTTGTCT GTAATTTTAC CAGAAAAGGC AAGAACCAAT TCAACTTGAC CAGCCTCAGC	3420
CAATTCGATA TGAAGGGCTT CATTGTCATG GTCAACTGTA AATGGACGAG CTTGACCTGC	3480
AACTTCTACA GAGGTGATTT CCAAATCTTT TTGGTGGAGG GAGATCCGGT CACTCTGTGC	3540
TTGACCAGTG ATGGTCACTT TCCCAGAAAA AGTCTTGCTC TCACGACTCA AATCTAAAAA	3600
TAAATCATAA TGTTCAGGAA CAAATTGCTT AATGGG	3636

(2) INFORMATION FOR SEQ ID NO: 79:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5066 base pairs

634

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

ATAGCGTGTA ATAATCGATT TTAGAGGTAC CATAAGCCAC CTCCTACAAA TAGAAACCGA	60
TATAAATCAA TGCCTTCCAC CCTTAGACTT CCCTAGTTCC TGTCTCAAGC GAAACATTTT	120
TTTGAAACAG GAATAAGTTA ACCAATTCAT ACCAATAGCT AGCAGAATAA AAAGAAACCA	180
AATGCCCCAT AACTTGATAT CTGTACATT TCTCAAGACG GTATTGAAAA ACAGAACTGA	240
AACAACCTGTC CAAGCAAGGC TAAAAAGAGA ATAGAAGGGG ATGTAAAACC AGTAAAAATA	300
ATAAAAAATT GGAAAAAACT TACTATTCTT GTTGGCCTTT TCAATCCAGT TATCAAAATA	360
AAAGTACGGT GCTAAAAGTA AGAATTTAAA CAAATGTTCC ATCACCGACA TCCCCCCTTC	420
TTTTGATAGC GTTTTCTATT ATTTTATTAT ATCAAAAAAA TCCGGAACGT TCATTCCAGA	480
TTCTACTTTT TTATTTGCGT TTTCTTGCGA TGAGATGAAT CGGTGTTCCC TCAAAAACAA	540
AGGCCTTGCG GATTTGATTT TCCAAGAAAC GCAGGTAAGA AAAGTGCATG AGTTCTTCTT	600
CATTGACAAA GATGACAAAG GTTGGTGGTT TGGTTGCCAC TTGGGTCGCA TAGAAAATCT	660
TGAGACGTTT TCCTTTGTCT GTCGGTGTG GGTGATGGC AATGGCATCC ATGATGACAT	720
CGTTCAAGAC AGCTGATGGA ATACGTGTAT TTTGACTTTC GCTGATTTGC TTAATCATCT	780
CAGGAAGTTT GTGGAGACCT TGCTTGTTA AAGCTGATAC AAAGATAATC GGTGCGTAAG	840
GCAGGTATTG GAACTGCTCA CGGATATCTT CTTCCCAGTT TTTCATAGTG TGGTTATCTT	900
TTTCAAGCGT ATCCCACTTG TTGACCACGA TAATCATCCC TTTACCAGCT TCATGGGCAA	960
ATCCTGCGAT ACGCTTGTG TACTCACGAA TGCCTTCTTC CGCATTGATG ACCATCAAGA	1020
CCACATCTGA ACGGTCAATA GCACGCATGG CACGCATAAC AGAGTATTTT TCAGTATTTT	1080
CATAAACCTT ACCAGACTTA CGCATACCAG CCGTATCAAT CATGGTAAAC TCTTGACCAT	1140
CTGTATCTGT AAAGTGGGTA TCAATGGCAT CACGAGTTGT TCCAGCAACA GGACTAGCAA	1200
TAACACGGTC TTCTCCCAAG ATAGCATTGA TCAAGCTTGA TTTTCCAACG TTAGGACGAC	1260
CAATCAAGCT AAACCTAATG ACATCTGGAT TTTCTTCCTC ATATTCATTT GGAAGATTTT	1320
CTACGATCGC ATCTAGCACA TCCCCTGTAC CGATTCCATG GACAGATGAG ATAGGCAATG	1380
GTTCACCCAA ACCGAGAGCA TAGAAATCAT ATATATCATT TCTCATCTCA GGGTTGTCCA	1440
CCTTGTGAC TCGGAGGATA ACTGGTTTGT GGGTCTTATA AAGCTTACGA GCTACGTATT	1500
CGTCTGCATC AGTAATTCCT TCCTTACCAG ACACGACAAA AACCATAACA TCTGCTTCTT	1560

635

CCATGGCAAT TTCTGCCTGG TGCTTGATTT GTTCCATGAA AGGAGCATCG ACATCATCAA	1620
TTCTCCTGT ATCAATCATG CTAAAAGAAC GATTGAGCCA CTCACCCGTT GCATAAATAC	1680
GGTCACGTGT CACTCCTTCG ACATCTTCTA CAATGGAGAT TCGCTCACCA GCGATCCGAT	1740
TAAATAGGGT TGATTTCCCA ACATTGGGAC GTCCTACAAT GGCAATAGTT GGTAGGGCCA	1800
TAATTTCTCA CTTTCTACAA TAATTTCTTC TGTTCAGAT TTTTCTAGT TGAGCTTGGT	1860
TCAGCTTGAC CAAACTGTTT TGCTAGGCGC TGACTCCAGC TTGTGGTCCG ACGCGCCCCA	1920
GCATAGTCAG CCTGAACACG GTCATAAGCT TGGATTGCCT CAGTTGACTG TTCTTGGTAT	1980
TCTTCCTCAA AGACAACATT CTCTAGTGGC AGTCTCGGTT TCATATCATG ATGTTGATTT	2040
GGCACACCCA GTGCCATCCC AAAGACAGAA TAGGTGTAGT CAGGTAGGTT AAAGAGCTCT	2100
GCCACTTCTT CAGACTTGTA TCGAACCAAA CCGATAATCA CACCACCATA GCCCAAGCTT	2160
TCAGCTGCCA ACAAGGCGTT TTGTCCAGCA AGAGCTGCAT CGACCGAACT AATCAAGAGA	2220
CCTTCCACAC CTTGGGGTTG GAAGGTGTCG GTATGAAGTC GGGCTCCCTT TTCTGCTCGG	2280
TTCAAATCTC CGACAAAGAG AAGGAAAACA GCAGACTGGC GAATGGCTTC TTGAGGTACC	2340
AATTCATACA AGGCATCTTT CTCTCTTGA CTTCGTACCA CAATCACAGA GTAGGATTGG	2400
AAATTCTTCC AAGATGATGC CATCTGGGCT GCTGTCAAAA TCTCATTTAA GTCTACTTGG	2460
GGAAATTTCTT GCTCTTTAAA CCTGCGCACT GAAGTATGAG CCTTCATCAA TTAAATGGTT	2520
TCTGTATCG ACGGTTTACT CCTTCTAAAC GAGTCTCCTC AGCCAAATAA CGGATGCGTT	2580
CCATGACCCG TCTGGCTTCC CAGGTTTCGT CATTTCCATG TTTCACTTTC GCAAAATGCT	2640
TCTCCAAATC TTCAAAGTTG AAGTTGGATG TGAAAAAGGT CGGTAAATTT TCCTGCATCC	2700
GATATTGGAG AATGACCTGC AGGATTTTCGT CACGCACCCA AACGGTTGAT TGCTCGGCGC	2760
CAATATCATC TAAAATCAGG ACCTCAGACA GCTTAATCTC ATCCACCAAG GTCTTAACAT	2820
TGCCATCACT GATAGCATTT TTGACATCAA TGACAAAGCT AGGATAGTGG AGGAGAGTTG	2880
ATGAAACACC ACGTTTTTCT GATAAATCAT GAGCTAAGCC CGCCACCATG AAACTTTAC	2940
CCACACCAAA GTCTCCATAT AAGTAAAGAC CTTTTCGAAT AGCTGGATAT TGCTCCACGA	3000
AGGCTAGTAG CTTTTCAAAA ACTGGTAAGC GCCCAAAATC ATCCAAGTCA ACTTGAGCCA	3060
AACTAGCTTT CTTGAGACTG GCTGGTAGAT TGATTAAGTT GAGACGGTTC TTAATAGCCG	3120
CTTCTTTTTC AGCCGCGATT AGCTCAGGAG TTTCTTCATA TGAAACATCT GCATAACCAT	3180
GATTCTTAAC CAAAATCGGC TTGTAGCCTT TGGCAATATA ATCCGTATCC CCACGGAGAA	3240
ACTTGTCACG CTCGGTGATG TACTGATTAA ACTTGAGAT ACTGCGATTT AATTCCTTTG	3300

636

GAGTTAAGGA TTCTTGCTGG ATAAAGGCCG CAACATCAGG GTCCTTCATG ATTTTCTGGA	3360
CCAAATCTTG ATAATAAAAA CGGCTGGGTT GACGTTTGAG TACGTCTCCG ACACTTTCCA	3420
TCTAATCTCC TCCTTTTCT AATCGAGCTA ATAGTTCTTG CTTCTTACGT TCTAGTTCCA	3480
GACGAGTTTC CTCGCTGGTT TCATTCTTAT ATTCAGGATT ACTCCATTTA GGAACATTGG	3540
TTTTTCTGG GGCAGTCTGA TTCTGTTTTT GTGTTTTTGC TTTCTGCCCT CGATCACGAA	3600
TTCGTAAAAC GGCCTCTTCT GCCGAATGAA TCTTTTGATA GGCATAGTCA TTGGCTACCT	3660
TCATGGCATA TTTCTATTG ATATTGCCG AATCCACCTT ATTAAGGTC AATAAGAGAA	3720
TAATATTGAT GACTTCGTCC AGTAAGCCCA AGCCAGCCAT CTGTTGCAAG AGTTCTCTTT	3780
CTGTTTGGGT AATGGTTCCC TTGCGTGTTT GCTTGATTTC TGCTAAGAAC TGCAGGGCAG	3840
TTTACTTTT AGCTTCTTTG ATAATGGTCG CTCCTTAAG ACTAAAGTCA GAGGAACTG	3900
GTTTTGAGC AATTTTTTCA CGCATGCGTT TGGTTGAAAT AACCTGGGAA ACAGCTGTTG	3960
ACTTGCCAA TTGATAGGTT TCAAACCAAG TCCATTTCTT CTCCTCGGCA ATAGCAAAGA	4020
GGTTTAAGAC ATCGGACTGC TCATCCGCAA AACGAAGTCC ATCTCGAGCC ATCAGCTGGC	4080
GAAAATGTTT CAAGTCAAAA TCATTGGCCA CTTTCTTCTT GAGACCAAGG TCTTCTTGAC	4140
TGCCTAGTTC TGCCAA'TCTT GGAAAGACTT GATTGAGTGA GACAGGTATT TCTTCACCAT	4200
CAGCACTTTC AACTTTCAAA TCCTCCACAG CTACATCGCC AATCTTTTTT TCTAAGAGTC	4260
TGCGATAAAC AGGATGCCCC AAGAAGTCTT GACTAGATAG AGGAGCATGG AGGGCTAGCT	4320
GATAAACATC ACCCTTTTGA TAGAGGGTCA AGAGATTAAA AGCAGATAAG ATTTTCAATG	4380
ATTTTATCAG TCTATCCATC CCAAAGTTGA GATGGTTGAG AATGCTTGAA AAAAGATATT	4440
CCTTTCTACC ATTATCCCAA AAAGTATTG TATAAAGATA AAGGCTCAGT GCCTCTGAC	4500
CGATAATCGG GAGGTAGCAC TGTACCAGAG ATGAGGTATC TTGCGACACC CGATTATTCT	4560
TTAGATAAGA AAAACGGTCA ATTGGCTTCA TTTATCTTTC CTTTTTCTTT TTAGAGGACT	4620
GGGTGATTG TTGGAGCAAG CTCTCTAACT CACTGACATC CTTAAACTA CGATAGACAC	4680
TAGCAAAACG TACATAGGTA ATCTCGTCCA ATTCAGCCAA CTCCTCCATG ACGAGTGAAC	4740
CAATGTCCTC ACTTTGAATT TCATTTCAT TTCGACCACG GAGTTTCTGT TCGATACGAT	4800
TGACTACCAT GTTGATTTC TCACTTGACA CAGGACGTTT CTGGGCTGAG CGGATAATCC	4860
CATTAAAGAT TTTATCTCTG GAGAATTGTT CCCGTGTGCC ATCTTTTTTA ACAACCACTA	4920
AGGTTCCTTC TTCTACTCGT TCGTAGGTTG TAAACGGTG TTGGCATTCT TCGCACTCAC	4980
GTCTTCTACG AATGGTGTTT CTTCTTCTG CTGGCGACT ATCGATAACA CTTGACTTGG	5040
TAGCCCCACA TTTTGGACAG GGTACC	5066

637

## (2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9607 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

CACCTGAAGT ATTGAAACA GCTATGGAAA ACATCATGCC TGTACTTGAA GTACGTGCAC	60
GTCGTGTTGG TGGTTCTAAC TACCAAGTCC CAGTTGAAGT TCGTCCAGAA CGTCGTACAA	120
CACCTGGACT TCGTTGGTTG GTAACAATCG CTCGTCTTCG TGGTGAACAC ACAATGCAAG	180
ACCGTCTTGC AAAAGAAATC TTGGATGCTG CTAACAACAC TGGTGCAGCA GTTAAGAAAC	240
GTGAAGATAC TCACCGTATG GCTGAAGCTA ACCCTGCATT CGCACACTTC CGTTGGTAAG	300
ATAGGATGCC AAACCGTTAA GAAAGTCCCA GAGAAAATAG GGAATCGAAG CAGGTTGCCG	360
TTGCAACCAA TGACATTCAT CTTTTCTCC AGACTTTTAG CTTGAGCTCA ACTAAATCAT	420
GATGCTAGGA ACGGTAAGGA TGCAAGGTAA AAATAGGAAA CTGACGCAGT ATTCCACGAA	480
TACAAGGAGT TTTATCTTTT TCACGCAGCA TCCCGTTCCA GCTCACATCG GCTAACTAAC	540
TTTAGCCCGG GTTCAAATTA GCTAAATCGA TTAGTATTAG CTATAACTCA GCTTACCATC	600
TCGTAAGTTG AAACCAACAA TAGCATGAAA ACATTGAGAA CCGGTAGGTC CTGCCTATCC	660
GTTTTTATTA AAATCGTGTT ATAATAGAAT AGAAATCAAA AATAAATAGG AGAAACAAAC	720
CTCATGGCAC GCGAATTTTC ACTTGAAAAA ACTCGTAATA TCGGTATCAT GGCTCACGTC	780
GATGCCGGTA AAACAACAAC TACTGAGCGT ATTCTTTACT AACTGGTAA AATCCACAAA	840
ATCGGTGAAA CTCACGAAGG TCGGTCACAA ATGGACTGGA TGGAGCAAGA GCAAGAACGT	900
GGTATCACGA TCACATCTGC TCGCAGCACA GCTCAATGGA ACAACCACCG CGTAAACATC	960
ATCGACACAC CAGGACACGT GGACTTCACA ATCGAAGTAC AACGTTCTCT TCGTGATTG	1020
GATGGTGCGG TTACCGTTCT TGAATCACA TCAGGTGTTG AGCCTCAAAC TGAACAGTT	1080
TGGCGTCAAG CAACTGAGTA CGGAGTTCCA CGTATCGTAT TTGCCAACAA AATGGACAAA	1140
ATCGGTGCTG ACTTCCTTTA CTCTGTAAGC ACACTTCACG ATCGTCTTCA AGCAAATGCA	1200
CACCCAATCC AATTGCCAAT CGGTTCTGAA GATGACTTCC GTGGTATCAT TGAATTGATC	1260
AAGATGAAAG CTGAAATCTA TACTAACGAC CTTGGTACGG ATATCCTTGA AGAAGACATC	1320
CCAGCTGAAT ACCTTGACCA AGCTCAAGAA TACCGTGAAG AATTGATTGA AGCAGTTGCT	1380

638

GAAACTGACG AAGAATTGAT GATGAAATAC CTCGAAGGTG AAGAAATCAC TAACGAAGAA	1440
TTGAAAGCTG GTATCCGTAA AGCGACTATC AACGTTGAAT TCTTCCCAGT ATTGTGTGGT	1500
TCAGCCTTCA AAAACAAAGG TGTTC AATTG ATGCTTGATG CGGTTATCGA CTACCTTCCA	1560
AGCCCACTTG ACATCCCAGC AATCAAAGGT ATTAACCCAG ATACAGACGC TGAAGAAATT	1620
CGTCCAGCAT CTGACGAAGA GCCATTTGCA GCTCTTGCCT TCAAGATCAT GACTGACCCA	1680
TTCGTAGGTC GTTTGACATT CTTCCTGTT TACTCAGGTG TTCTTCAATC AGGTTTCATAC	1740
GTATTGAATA CTCTTAAAGG TAAACGTGAA CGTATCGGAC GTATCCTTCA AATGCACGCT	1800
AACAGCCGTC AAGAAATCGA CACTGTTTAC TCAGGTGATA TCGCTGCTGC CGTTGGTTTG	1860
AAAGATACTA CAACTGGTGA CTCATTGACA GATGAAAAAG CTAAATCAT CCTTGAGTCA	1920
ATCAACGTTT CAGAACCAGT TATCCAATTG ATGGTTGAGC CAAAATCTAA AGCTGACCAA	1980
GACAAGATGG GTATCGCCCT TCAAAAATTG GCTGAAGAAG ATCCAACATT CCGCGTTGAA	2040
ACAAACGTTG AAAGTGGTGA AACAGTTATC TCAGGTATGG GTGAACTTCA CCTTGACGTC	2100
CTTGTTGATC GTATGCGTCG TGAGTTCAAA GTTGAAGCGA ACGTAGGTGC TCCTCAAGTA	2160
TCTTACCGTG AAACATTCCG CGCTTCTACT CAAGCACGTG GATTCTTCAA ACGTCAGTCT	2220
GGTGGTAAAG GTCAATTCCG TGATGTATGG ATTGAATTTA CTCCAAACGA AGAAGGTAAA	2280
GGATTGGAAT TCGAAAACGC AATCGTCGGT GGTGTGGTTC CTCGTGAATT TATCCACGG	2340
GTGAAAAAG GTTTGGTAGA ATCTATGGCT AACGGTGTTC TTGCAGGTTA CCCAATGGTT	2400
GACGTAAAG CTAAAGCTTTA TGATGGTTCA TATCAGGATG TCGACTCATC TGAAACTGCC	2460
TTCAAGATTG CGGCTTCACT TTCCCTTAAA GAAGCTGCTA AATCAGCACA ACCAGCTATC	2520
CTTGAACCAA TGATGCTTGT AACAATCACT GTTCCAGAAG AAAACCTTGG TGATGTTATG	2580
GGTCACGTAA CTGCTCGTCG TGGACGTGTA GATGGTATGG AAGCACACGG TAACAGCCAA	2640
ATCGTTTCGTG CTTACGTTCC ACTTGCTGAA ATGTTCCGTT ACGCAACAGT TCTTCGTTCT	2700
GCATCTCAAG GACGTGGTAC ATTCATGATG GTATTTGACC ACTACGAAGA TGTACCTAAG	2760
TCAGTACAAG AAGAAATTAT TAAGAAAAAT AAAGGTGAAG ACTAATCCGT CCTCACTCTA	2820
GAAGGAAGTC ACTTAGTGGC TTCTTTTGT CTTTAGAAAA TACCTCTAAA TATGGTAAAA	2880
TAGTAGAAGA ATAATGTGAG GAAATGAAT GTCAAATAGT TTTGAAATTT TGATGAATCA	2940
ATTGGGGATG CCTGCTGAAA TGAGACAGGC TCCTGCTTTA GCACAGGCCA ATATTGAGCG	3000
AGTTGTGGTT CATAAAATTA GTAAGGTATG GGAGTTTCAT TTCGTATTTT CTAATATTTT	3060
ACCGATTGAA ATCTTTTATG AATTAAAGAA AGGTTTGAGC GAAGAATTTT CTAAGACAGG	3120
CAATAAGCT GTTTTGAAG TTAAGGCTCG GTCTCAAGAA TTTTCAAATC AGCTCTTGCA	3180

639

GTCCTACTAT AGGGAGGCTT TCTCTGAAGG TCCATGTGCT AGTCAAGGTT TTAAGTCCCT	3240
TTATCAAAAT TTGCAAGTTC GTGCTGAGGG TAATCAGCTA TTTATTGAAG GATCTGAAGC	3300
GATTGATAAG GAACATTTTA AGAAGAATCA TCTTCCTAAT TTAGCCAAAC AACTTGAAAA	3360
GTTTGGTTTT CCAACTTTTA ACTGTCAAGT CGAGAAGAAT GATGTCCTGA CCCAAGAGCA	3420
GGAAGAGGCC TTTCATGCTG AAAATGAGCA GATTGTTCAA GCTGCCAATG AGGAAGCGCT	3480
CCGTGCTATG GAACAACTGG AGCAGATGGC ACCTCCTCCA GCGGAAGAGA AACCAGCCTT	3540
TGATTTTCAA GCGAAAAAAG CTGCAGCTAA ACCCAAGCTG GATAAGGCGG AGATTACTCC	3600
TATGATCGAA GTGACGACAG AGGAAAATCG TCTGGTATTT GAAGGGGTTG TTTTGTATGT	3660
GGAGCAAAAA GTGACTAGAA CAGGTCGTGT TTAAATCAAC TTTAAATGA CGGACTATAC	3720
TTCAAGTTTT TCTATGCAA AGTGGGTTAA AAACGAGGAA GAGGCCCAGA AGTTTGACCT	3780
CATCAAGAAG AATTCTTGGC TCCGAGTTCG AGGGAATGTG GAGATGAATA ACTTCACACG	3840
CGATTTGACT ATGAACGTAC AGGATCTGCA GGAAGTTGTT CACTATGAGC GGAAGGATTT	3900
GATGCCAGAA GGTGAGCGTC GGGTTGAGTT TCATGCTCAT ACTAACATGT CGACTATGGA	3960
TGCTTTGCCA GAGGTGGAAG AGATTGTTGC AACAGCTGCT AAGTGGGGAC ACAAGGCGGT	4020
TGCTATCACG GACCATGGGA ATGTCCAGTC CTTTCCACAT GGCTATAAGG CGGCTAAGAA	4080
AGCGGGAATC CAGCTGATCT ATGGGATGGA AGCCAATATC GTGGAGGACC GTGTCCCTAT	4140
CGTCTATAAC GAAGTGGAGA TGGACTTGTC AGAAGCAACC TACGTGGTCT TTGACGTGGA	4200
AACGACGGA CTTTCAGCTA TCTATAATGA CTTGATTCAG GTTCCGGCTT CTAAGATGTA	4260
CAAGGGGAAT GTTATTGCTG AATTTGATGA ATTTATCAAT CCTGGGCATC CCTTGTCAGC	4320
CTTTACTACA GAGTTAACTG GAATTACAGA TGATCATGTC AAAAATGCCA AACCCTAGTA	4380
ACAAGTTTTG CAAGAATTCC AAGAATTTG CAAGGATACG GTCCTAGTTG CCCACAATGC	4440
TACCTTTGAC GTTGGCTTTA TGAATGCTAA TTATGAGCGG CATGATCTTC CAAAGATTAG	4500
TCAGCCAGTT ATTGATACGC TGGAGTTTGC TAGAAACCTC TATCCTGAGT ATAAACGCCA	4560
TGGTTTGGGG CCTTTGACCA AGCGTTTTGG TGTGGCCTTG GAACATCACC ACATGGCCAA	4620
CTACGATGCG GAAGCGACTG GTCGTCTGCT TTTCATCTTT ATCAAAGAGG TAGCAGAAAA	4680
ACATGGTGTG ACCGATTTAG CTAGACTCAA CATTGATCTA ATCAGTCCAG ATTCTTACAA	4740
AAAAGCTCGG ATCAAGCATG CGACCATCTA TGTCAAGAAT CAGGTAGGTC TAAAAAATAT	4800
CTTTAAGCTG GTTTCCTTGT CTAATACCAA GTATTTTGAA GGAGTGCCAC GGATTCCGAG	4860
AACGGTTCTA GATGCCCATC GAGAGGGCTT GATTTTAGGT TCAGCCTGTT CAGAGGGTGA	4920

640

AGTTTTTGAC	GTGGTCGTTT	CTCAAGGTGT	GGATGCGGCG	GTGAGGTGG	CCAAGTATTA	4980
TGATTTTATC	GAGGTCATGC	CACCGGCTAT	CTATGCACCC	TTGATTGCCA	AAGAGCAGGT	5040
CAAGGATATG	GAGGAACTCC	AGACCATTAT	CAAGAGTTTG	ATAGAGGTTG	GAGACCGCCT	5100
TGGCAAGCCT	GTTCTGGCTA	CGGAAATGT	TCACTATATC	GAACCGGAAG	AAGAGATTTA	5160
TCGTGAAATT	ATCGTCCGTA	GTTTGGGACA	GGGTGCGATG	ATTAATCGAA	CTATCGGTCA	5220
TGGTGAACAT	GCCCAACCAG	CACCACTTCC	AAAGGCTCAT	TTTCGAACGA	CTAATGAGAT	5280
GTTGGATGAA	TTTGCCTTTT	TGGGAGAGGA	ACTGGCTCGT	AAACTGGTTA	TTGAAAACAC	5340
CAATGCCTTG	GCAGAAATAT	TTGAATCCGT	TGAAGTCGTT	AAGGGTGAAT	TGTATACGCC	5400
TTTCATCGAC	AAGGCTGAAG	AAACAGTTGC	TGAGTTGACC	TATAAGAAAG	CTTTTGAGAT	5460
TTATGGAAT	CCGCTGCCAG	ATATTGTTGA	TTTGGCGATT	GAAAAAGAAT	TAACATCCAT	5520
ACTGGGGAAT	GGATTGCTG	TGATTTATCT	GGCATCGCAG	ATGCTGGTGC	AACGTTCTAA	5580
TGAACGGGGT	TATTTGGTTG	GTTCTCGTGG	GTCTGTCCGA	TCTAGTTTCG	TTGCGACCAT	5640
GATTGGGATT	ACGGAGGTCA	ATCCTCTCTC	TCCTCACTAT	GTCTGTGGTC	AGTGTCACTA	5700
CAGTGAGTTT	ATCACAGATG	GTTCTGACGG	TTCAGGATTT	GATATGCCCC	ATAAGGACTG	5760
TCCAAACTGT	GGTCACAAAC	TCAGTAAAAA	CGGACAGGAT	ATTCCGTTTG	AGACCTTCCT	5820
TGCTTTTGAT	GGGGATAAGG	TTCTTGATAT	TGACTTGAAC	TTCTCGGGAG	AAGATCAGCC	5880
TAGCGCCAC	TTGGATGTGC	GTGATATCTT	TGGTGAAGAA	TATGCCTTCC	GTGCGGGAAC	5940
GTTTGGTACG	GTAGCTGCCA	AGACTGCCA	TGGATTGTGC	AAAGGTTACG	AGCGAGATTA	6000
TGGCAAGTTT	TATCGTGATG	CAGAACTAGA	ACGCCTCGCT	CAAGGAGCGG	CGGGTGTCAA	6060
GCGGACAACA	GGCCAACACC	CGGGGGGAAT	CGTTGTTATT	CCGAACTACA	TGGATGTCTA	6120
CGATTTTACG	CCTGTCCAGT	ATCCAGCAGA	TGATGTCACG	GCTGAATGGC	AGACCACTCA	6180
CTTTAACTTC	CACGATATCG	ATGAGAACGT	CCTCAAACCT	GATGTACTGG	GACATGATGA	6240
TCCGACTATG	ATTCGAAAAC	TTCAGGATTT	GTCTGGTATT	GACCCTAATA	AAATTCCTAT	6300
GGATGACGAA	GGCGTGATGG	CACCTCTTTC	TGGGACTGAT	GTGCTAGGGG	TAACACCTGA	6360
ACAAATTGGA	ACGCCTACGG	GTATGTTGGG	GATTCCAGAG	TTTGGAAACAA	ATTTCTGTACG	6420
TGGAATGGTA	GACGAAACCC	ATCCGACAAC	CTTTGCGGAA	TTGCTTCAGC	TGTCTGGTCT	6480
GTCCACCGGT	ACTGATGTTT	GTTTGGGGAA	TGCTCAGGAT	CTGATTAAGC	AAGGAATAGC	6540
GGACCTATCG	ACTGTTATCG	GTTGTCCGGA	CGACATCATG	GTTTACCTCA	TGCATGCGGG	6600
TCTGGAACCT	AAGATGGCCT	TTACCATTAT	GGAACGGGTA	CGTAAGGGTT	TGTGGCTAAA	6660
GATTTACAGAA	GAGGAGAGAA	ATGGCTATAT	CGAAGCAATG	AAGGCTAATA	AGGTGCCAGA	6720



641

GTGGTATATC GAATCCTGTG GGAAAATTAA GTACATGTTT CCTAAGGCCC ATGCCGCAGC	6780
CTACGTTATG ATGGCCTTGC GTGTAGCTTA CTTCAGGTT CACCATCCTA TTTATTACTA	6840
CTGTGCTTAC TTCTCCATTC GTGCTAAGGC TTTTGATATC AAGACCATGG GTGCCGGCTT	6900
GGAGGTCATC AAGCGCAGAA TGGGAAGAAAT CTCTGAAAAA CGGAAGAACA ATGAAGCCTC	6960
TAATGTGGAA ATCGATCTCT ATACAACCTT TGAGATTGTC AATGAGATGT GGGAACGAGG	7020
TTTCAAGTTT GGTAAATTAG ATCTCTACTG TAGTCAGGCG ACAGAGTTCC TCATCGACGG	7080
GGATACCCCTT ATCCCACCAT TTGTAGCAAT GGATGGTCTG GGAGAGAACG TTGCCAAGCA	7140
ACTGGTGCGG GCGCGTGAAG AGGGAGAATT CCTCTCTAAA ACAGAACTAC GCAAGCGTGG	7200
TGGACTCTCA TCAACCTTGG TTGAAAAGAT GGATGAGATG GGTATTCTTG GAAATATGCC	7260
AGAGGATAAC CAGTTGAGTT TGTGTGATGA GTTGTTTTAA AAAATTGCTT AATAATCTAT	7320
TAAAAGAGGC TAACGTATAT CCAATAGATT TACATTAGCT TTCTTTTTTG TTAATAAGT	7380
CTATGGAAAG AGGGTGAGAG TATGTCAAAG ATGAGTATAA GCATCCGTCT GGATAGTGAG	7440
GTTAAGGAGC AGGCCCAACA GGTGTTTAGT AATCTGGGAA TGGATATGAC AACAGCTATT	7500
AATATTTTCC TTCGTCAGGC AATTCAATAT CAGGGATTAC CTTTTGATGT TAGACTAGAC	7560
GAAAATCGGA AGTTGCTCCA AGCGTTAACG GATTTAGACC AAAATCGTAA TATGAGCCAG	7620
TCTTTTGAAT CAGTCTCAGA TTTGATGGAG GACTTACGTG CTTAAGATTC GTTATCATAA	7680
ACAGTTTAAA AAAGATTTTA AGTTGGCTAT GAAGCGTGGT TTGAAGGCAG AATTATTAGA	7740
AGAAGTTTGG AATTTTCTGG TTCAAGAAAA AGAACATCCT GCCAGAAATC GTGATCATTC	7800
ATGACGGCA TCCAAGCATT TTCAAGGAGT TCGTGAATGC CATACCCAGC CAGATTGGCT	7860
TTTGGTTTAT AAAGTAGACA AGTCGGAATT GATTTTAAAT TTGCTGAGGA CAGGCAGTCA	7920
CAGTGATTTA TTTTAATCTA TTTTAAGGGG GTTCTCATGA AACTAAGAAT ATTTGCGGAA	7980
GATAAGCCGG CTAAGAAGGT ATTTGAATAT CAATTAGAAC TTGCTGATCG TACAATTCTT	8040
CTATCGACAG CACTCTTCTC AGGTGCTATT GCTTTAGCAG GAATCTTTTC TGCTTTGAAA	8100
GAAAAATAAA AATAGAAAAG AGAAAACAGA ATGGTTTTAC CAAATTTTAA AGAAAATCTA	8160
GAAAAATATG CGAAATTGTT GGTGCGAAC GGAATTAACG TGCAACCTGG TCACACTTGT	8220
GCTCTCTCTA TTGATGTGGA GCAACGTGAA TTGGCACATC TAATCGTGAA AGAAGCTTAT	8280
GCCTTGGGTG CGCATGAGGT CATCGTTCAG TGGACAGATG ATGTGATTAA CCGTGAGAAA	8340
TTCTCTCATG CCCCAGTGA GCGTTTGGAC AATGTGCCAG AATACAAGAT TGCTGAGATG	8400
AACTATCTCT TGGAGAATAA GGCTAGCCGT CTGGAGTTC GTTCATCTGA TCCAGGTGCC	8460

642

TTGAACGGAG TGGACGCTGA CAAGCTTTCA GCTTCTGCTA AAGCTATGGG ACTTGCCATG	8520
AAGCCTATGC GTATCGCAAC TCAATCTAAC AAGGTTAGCT GGACTGTAGC AGCTGCAGCA	8580
GGACTTGAGT GGGCTAAGAA AGTCTTCCCA AATGCTCCGA GCGACGAAGA AGCAGTTGAT	8640
TTCCTTTGGG ACCAAATTTT CAAAACCTGC CGTGTCTACG AAGCAGATCC TGTTAAGGCT	8700
TGGGAGGAAC ATGCAGCCAT TCTCAAGAGC AAGGCCGATA TGCTTAATAA GGAGCAATTT	8760
TCAGCCCTTC ACTACACAGC GCCAGGAACA GATTTAACAC TTGGTTTGCC AAAGAACCAC	8820
GTTTGGGAAT CAGCTGGTGC TGTCAATGCA CAGGGCGAAG AATTCTTGCC AAATATGCCA	8880
ACAGAAGAGG TCTTCACAGC GCCTGACTTC CGTCGTGCAG ATGGTTATGT CACTTCTACA	8940
AAACCGCTTA GCTACAACGG AAATATCATT GAAGGCATTA AGGTGACCTT TAAGGATGGA	9000
CAAATCGTAG ATATCACTGC TGAGAAGGGT GATCAGGTTA TGAAAGACCT TGTCTTTGAA	9060
AATGCGGGTG CGCGTGCCCT GGGTGAATGT GCCTTGGTAC CAGATCCAAG TCCAATTTCT	9120
CAGTCAGGCA TTACCTTCTT TAACACCCTT TTCGATGAAA ATGCGTCAAA CCACTTGGCT	9180
ATCGGTGCAG CCTATGCGAC TAGCGTTGTT GATGGAGCGG AGATGAGCGA AGAGGAGCTT	9240
GAAGCTGCAG GGCTTAACCG TTCAGATGTT CACGTAGACT TTATGATTGG TTCTAACCAA	9300
ATGGATATCG ATGGTATTCG TGAGGATGGA ACGCGGGTAC CTCTTTTCCG TAATGGGAAT	9360
TGGGCAAATT AAGGAGATAA TATGTTAGGA AGTATGTTTCG TTGGTCTCCT AGTGGGATTT	9420
TTAGCAGGTG CTATGACCAA TCGTGGAGAG CGAATGGGAT GTTTTGGAAT AATGTTTCTC	9480
GGTTGGATCG GAGCCTTTCT AGGTCACCTG CTCTTTTGAA CTTGGGGGCC AGTTTATCA	9540
GGAACAGCTA TTATCCCAGC GATTTTAGGA GCCATGATTG TTTTAGCTAT TTTTGGAGA	9600
CGAGGAA	9607

## (2) INFORMATION FOR SEQ ID NO: 81:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14231 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

CTACAAGATA ATTCCAGCTA TAACATCCGC TATAATAGTA AGAGCGAGCT CTATGATAAG	60
GCTCATTAGT TTCACCTCCT CTCACGAACC CATAGGAACG TAATCGGTAA CCGATGACAA	120
AAATAGTATA CCACAATACA TTTAGATCAT CAAGGTCAC TAAATCTTGA AATATCAGAT	180
CTAAGAGAAA AATCTTTAAA ATCAGAAAAA CGCATAATAT CAGGTGTGCA AAAACTTGAT	240

643

ACTATGCGTT TTATTGTGGG AAGGTTTACT CCATTTTCTC CTGAAATTGA GTTTTGTGCC	300
AGCCTCTGTT TTTAGGGTTG CTAAGAAAAT AATGTCATGT GGTGAATATT TGTAAATCAG	360
TCAGCAGACA GAACGATACT CTTCGAAAAT CTCTTCACAT CATGTCAGCT TCGTCTTTCC	420
GTATATATGT GACTGACTTC ATCAGTTCTA TCTACAACCT CAAAACAGTG TTTCGAGCTG	480
ACTTGATCAA TTTTCAAATC TGTACTTTGA GCAAGCTGAG ACTAGCTTCC TATTTGATTT	540
TCATTGAATA TCAGAAACCC ATTCTCCATC AAATAATTCTG ACTGCGTCTA ATAATTTTGT	600
ATCTGGCAGC GTGTCTGAAA TAAAGGTTGT GTATTTGGAG AGGGGATTAA TTTTAAAAAA	660
TCCAGTCTTG TAAAATTTAG AACTATCAAT CAGTAAGATG GTTTCATGGG CTTTGTCAAT	720
AATATTCTTT TTTGAAATAG CTTGGCTGAG AGAAGCTTCA TAAACATATT GGTCAATCAAT	780
ACCTCTTGCT GAACAAAATG CTAAATCGAT ATTAATAATGA TCTAATAAAG AATTTTCCTT	840
ATCATAGTTG ACCACGGAAC AGGATTGATG TTTGACCTCG CCAGATGTGA TAAAGATTTT	900
GGAGCTATCT TTAACAGTTT CAGATAGGGT TTGTGCAGTA TGTAAACCAT TTGTAAAAAT	960
AATCAAATTA TCAAGTTCAG AAAGATAGGG ACAGAGTTCC TAGACAGTAG TACTAGAATC	1020
TAGATAGATA CACATACCAG ACCGAATAAA GTCTTTAGCG AGACTAGCGA TTAGTCTTTT	1080
TTGCCTAGTA CTTTCTCCTT CACGTATTTG ATGAGAAAGT TCAATTGTGT TCATAGAGGA	1140
CAGGGTCACG TATCCGTGCT TTCTTTTGAT AAGACCTTGA TTTTCTAAGA AAATTAAATC	1200
ACGACGTAAG GTAATTGTGC TGGAGAAAGT GATTTCTGCC AGCTCTTTTA CGGCAATTCT	1260
TTTTTCTTTT TTGATAATTT CAATCAATTC AAGTACACGT TCATCTTTTA TCATAAGCTC	1320
CTCCTAATTT ATCATTTCAA CTATATTATA GCACAAATTG GAGGAATTTG AATTATTTTT	1380
ATGAATATTG GGTTAACATT TGAACATTAT TCAAGTAAGC GTTCACATAT TGAAAAAATA	1440
AAACGTGGGG ATTATAATAA AGTTAATCma GGACGAAGAG AGAAGAAAAA TGGAAGCGGT	1500
TTTAGCAATA GATTTAGGTG CGACTTCTGG AAGAGCAATC GTTGGTTACC TTTCTGAAAA	1560
TAAACTAGTA ATGGAAGAAA TAAATCGCTT TTCTAATCTA CCTATTAGAG TAAAAGGGCA	1620
TTTATCTTGG GATATTGACT TTCTACTAGC TAAAATTCTT GAAAGTATCC GCTTGGCTAA	1680
TACTAGTTAC AAGATTTTAT CTATCGGTAT TGACACATGG GGAGTTGATT TTGGACTGAT	1740
TGATAATGAA GGTAAGCTGT TATTACAACC TGTTCATTAT CGTGATGAAA GAACAAAGGG	1800
AGTGTTAAAG GAAATATCTG AAATGACTGA ATTAGAAAAA CTGTATTCTAG AGACAGGAAA	1860
TCAGATTATG GAGATAAATA CCTTGTTCCTA ACTCTTTAAG GCACGTCAAG AATCTCCTGA	1920
CTCTTTCTAT AAGACCAATA AGATTCTTTT AATGCCAGAT TTGTTAATT ATCTCTTGAC	1980

644

AGGTAAGTTT GCTACAGAAA AAAGCATTGC TTCAACAACCT CAATTATTTG ATCCTAGGAG	2040
TCAAAATTGG AATCAGAATA TCTTAAAACCT ATTTGAATTG GATTCACTCT TACTTCCTGA	2100
AATTGTTTCA GAGGGAAATG TTCTTGGAAG GATAAAAGAG GAGTATGGTT TAGGCGATAT	2160
TCCTGTTGTG AATGTTTGTA GTCATGATAC AGCAAGCGCG ATTGTCTCAG TACCTAAGAC	2220
AGAAGGTAGT TTATTTATTT CATCAGGTAC TTGGTCTTTG GTTGGAGTGG AACTTACTTC	2280
ACCGATTCTT ACTACCGAAT CCTTCAGTTA TGGATTTACA AATGAAGTCG GTAAAGATGG	2340
AGTGATTACA TTTCTGAAGA ATTGTACAGG GTTGTGGATC ATAGAGGAAC TAAGACGTTT	2400
ATTTGAACGA AGAGGGAAAG CCTATTCTTT TGATGATATT AGGACAATGG TGGAGAAAGA	2460
AAAAGAAAAT CTTCTCTGA TTGATACTGA ATCAACTGAA TTTGCAACAG AATCTGATAT	2520
GCACAAGACT TTGACAGAAT ATCTAGCTTA TCATCATGAA ACTAGAGACT GGACAGATGG	2580
ACAACTATTT AAGATTGTTT ATGAAAGCCT AGCTGAAACG TATAGGAAAG CGATAGAGTT	2640
ACTAGAAGAA CTAACCTATA AGGTTTATAA GAGGATATAT GTGATTGGAG GAGGTGCTAG	2700
AGCCAGTTAC TTTAACCAAA TGATTGCTGA TAGAACTGGT AAAGAGGTTC TTACAGGTTT	2760
GACTGAGGCT ACAGCTGTGG GGAATATTGT TGTGCAGCTC ATAGCTATGG GACAATTAAA	2820
AGGGATGGAA GAGGCTCACC ATGTTATTGA GGAGTTTCTA CAATTAGAGA GTTATTACTC	2880
CCAAAAGAAT TAAAAAGATT GAGAGTTTGT AAATTTGCCT CCTCCCCCT TCTTAGCTTT	2940
TGTGCAGGAA GGGGGGATAA TTGGTGAATT GAAAAATATT TAGTGTTTTG ATATGAGGAG	3000
GACAAGGATG TCAGATGTAA AACAAGAATT AATTAAATAT GGTAAGAAGC TAGTAGAAAC	3060
AGATTTGACG AAAGGAACAG GTGGGAATCT CAGCGTTTTT GATCGTGAAA AACAATTGAT	3120
GGCAATTACC CCGTCGGGTA TTGATTTCTT TGAATCAAA GAATCCGATA TTGTAGTGAT	3180
GGATATTAAT GGAAATGTTG TAGAGGGAGA ACGCTTGCCA TCTAGCGAAT GGTATATGCA	3240
TTTGATTCAA TATCAAACCTC GTGATGATAT CGATGCAATT ATCCATGCTC ATACAACCTA	3300
TGCAACAGTA TTAGCTTGTC TCAGAGAACC ACTTCCAGCG AGTCATTATA TGATTGCAGT	3360
GGCAGGGAAA GATGTTCCGG TAGCTGAGTA TGCAACATAT GGCACGAAAG AATTGGCTGT	3420
GAATGCAGCT AAAGCAATGG AAGGTCGTAG AGCAGTTTTA CTAGCGAATC ATGGAATTTT	3480
AGCAGGTGCA CAAAATTTAT TGAATGCATT TAATATTGTT GAAGAAGTTG AATATTGTGC	3540
AAAAATTTAT TGTTTAGCTA AGAATTTTGG AGAGCCAGTA GTTCTTCCTG ATGAGGAGAT	3600
GGAAATTGATG GCAGAAAAAT TAAAAACATA CGGTCAGAGA AAATAGGGAG GATATTAATG	3660
TTAAACATA TACCGAAAAA TATTTCTCCA GATTTATTGA AGACTTTAAT GGAAATGGGA	3720
CATGGAGATG AAATAGTATT AGCTGACGCG AATTATCCTT CTGCCTCATG TGCAAATAAG	3780

645

CTAATTCGTT GTGATGGTGT AAATATTCCA GAATTATTAG ATTCCATTCT GTATTTAATG	3840
CCATTAGATA GTTACGTCGA TAGTTCAATT CAGTTTATGA ACGTTGTTTC GGGTGATGAT	3900
ATTCCTAAGA TATGGGGTAC CTATAGACAG ATGATTGAAG GTCATGGTAC AGATCTTAAA	3960
ACGATTACTT ATCTTAGAAG AGAAGACTTT TATGAACGTA GTAAGAAAGC TTATGCTATT	4020
GTTGCTACAG GAGAACTTC ACTTTATGCT AATATTATCC TTAAGAAAGG AGTAGTTGTT	4080
GAAAGAGAAA ATGTTCAATA GAGGAATTTT AGTTGCCAGT CATGGTAATT TTGCTAGCGG	4140
AGCTCTCATG ACCGCAGAAA TGTGTTGTTGG TGAGACAACA AATGATAGAG TTAGGACATT	4200
AGGTTTGATG CCTGGAGAGA ATATTGTAGA GTTTGAGCAT TATTTTAAAA ATCAAGTGGA	4260
TGAAGTGTGA GACTCAAATC AAGAGGTTAT CGTTTGGACT GACTTGATTG GAGGAAGTCC	4320
TAATAATGTG GCTTTGTCAC GGTTTTTAAA TTTGGATTCA GTTGATATTG TAACAGGGTT	4380
TAATATCCCT CTCCTAGTGG AATTAATATC AAGTTATGAT TCAAAAATCA ATTTAGAAGA	4440
AATGTTCAC AATGCTCAAA ATAGTTTGTT TAATGTTAAA CAACAACTTA ACGTAGAGGA	4500
GGAAGAAGAT TTATGTCTAT AGAGTTTGTT CGTATTGATG ACCGTCTGGT ACATGGTCAA	4560
GTTGCTACTA CGTGGCTAAA AAAGTATGAT ATTGAGCAAG TTATCATTGT TAATGATCGC	4620
ATCTCAGAAG ATAAACACG ACAATCTATT TTAAAGATT CTGCACCGGT AGGTTTAAAA	4680
ATTGTTTTCT TTAGTGTAAG ACGGTTTGTT GAAGTTTAA ACTCTGTGCC AATAAAAAAG	4740
AGAACAATGC TGATATATAC AAATCCAAAA GATGTGTATG ATTCTATTGA AGGAAATTTA	4800
AAATTCGAGT ACCTCAATGT AGGACAGATG ACTAAACGG AGGAAAATGA AAAGGTAACG	4860
GGAGGTGTAG CTCTAGGTGA AGAAGACAAA TATTATTTTA AGAAAATAGT TGATAAGGGA	4920
ACGAGAGTTG AAATTCAAAT GGTTCCTAAT GATAAAGTTA CAATGTTGGA AAAATTTTTA	4980
TAAAAATAAT TTAAGGAGGT ACAGTATATG CTATTCACAC AAGCATTACT GGTGACATTA	5040
GTTGGGATTA TTGCCACTAT TGAATAAAT GGACCGTTAT TTATGATTCA CCGTCCGTTA	5100
GTTACAAGTG CAATGGTTGG CTTAGTATTA GGAGATTTC CCAAGGTGT TCTTATTGGT	5160
TCAGCTCTTG AATTAACTTG GCTCGGTGTA ACAGGTATTG GAGGTTATAC TCCACCAGAT	5220
ACTATTTTCT GTGCGATTAT TGGTACTGCA TTTGGTATTT TATCTGGTCA AGGAGAACT	5280
GCTGGTATCG CTATAGCAGT TCCAATTGCA GTTGCTACCC AACAGTTGGA TGTCTTGCA	5340
AAAATTTTAG ATGTTTATTT TGTGAAAAA GCTGATAATG ATGCTAAAA CGGAGATTAT	5400
TCAAAGATCG GTTTTTATCA TTATTCAAGT TTGGTTTTAA TCACGTTATT TAAAAATGTA	5460
CCAAATTTCC TAGCTATTAT GCTTGGAGGG GAATATGTGG CAGACTTCTT TGCTAAGGTT	5520

646

CCACCAATCG TTATGCAGGG ACTTAACTCT GCAGGTGCTT TACTACCTTC AATTGGTTTT	5580
GGTATGCTTT TAAATATGAT GCTCAAGAAA AATATGTGGG TATTCTTGTT GATTGGATTC	5640
ATTTGTTCTG TGTATGGAGG AATGTCAACC ATTGGGATCT CACTAGTTGG TATTGCGGTA	5700
GCATACTTCT ACGATATGAT TGGAAGCAAA CCACAAGAAA CAACTTCAAG TAGTGATGTT	5760
GAGGAGGATC TTGATCTATG ATGAATAATA AAGTAACTAA AGTTGAACTT AAAAAAGTTT	5820
TCAAACGAAG TTTTATGTAT GGTTCCTCAT GGAAGTATGA GAGAATGCAG AACCTAGGTT	5880
TTCTATATAC AATTCTTCCA GTATTGAAAA AACTATACCC AGACAAAGAT TCAGCTTCTC	5940
CTGCAATGAA ACGTCACCTT GAGTTTTCAT ATACTCATCA AACAGCGGCA CCATTTATTC	6000
TTGGAGTTAC TTCCGCTATG GAAGAACAAG AAGGAAATGA AGGTGCAGCT TCAATTACTG	6060
GTATTAAAGT TGGCTTGATG GGGCCACTGG CTGGTCTAGG AGATAGTTTG TTCTGGCTGA	6120
CACTAGTTCC TATCTGTTTT AGTATTGGTG CGTCTTATTC TAAAGACGGC GGTGCTTTAG	6180
GTATCTTTAT CGCCTTAATA TTGTTTAATA TTATTAATAT TCCTGTAAAA TATTCGGTT	6240
TGAAATATGG GTATACTAAG GGTTCCTAGT TTATCCAAGA AAATAATACA AAAGGAACAT	6300
TGAATCGCGT TACGAGTATG GCGACAGCAT TAGGGCTAGT ACTAGTGGGT GGTTCGATTC	6360
CATCAATCGT TGGTATTAAT TTTGGATTAG AATTTAAGCA GGGGGAAGTT GTTATTTCTG	6420
TTCAAGAAAT GATTACAAAA TTAATTCAG GATTTATCCC TATGGCTTTG ACTTTATTAA	6480
TGTGTAAATT AATTAGAAAA GGAAAGAATC CGGTTGTAAT AATCTTTAGT GTTATGGCTA	6540
TTGGAGTTAT TCTAGTTGTT TTAGGAATTT TGAAGTAGTA GAAAGTGTGG AGGTGGTATT	6600
TGGGATATCA CCTCCATTTT GGAAGAGAGG TAAAGAGTGA AATTATGGTA TAAGAAAGCT	6660
GGCGCAAATT GGAATGAAGC CTGCGCGATT GGGAACGGTC ATTTAGGTGG TATGATTTAT	6720
GGTTCAGCTA CAAAAGAAATG TATTCAACTA AACGATGAGA CTATTTGGTA TAGAGGAAAG	6780
TCAGATAGAA ATAATCCAGA CTCACTATTG CATCTTAAAA AAATTCGGGA ATATCTTTTA	6840
GATGGAGAAA TTCAGAAAGC CGAAGAATTG ATAAAGTTAA CAGTGTTTGC TACCCCAAGA	6900
GATCAAAGCC ACTATGAATT ACTTGGGGAA CTTTACATTG AGCATATAGA TATTCAGTCT	6960
TGTGCTCTTT CATTGTATGA AAGAGAGCTA GATTTAGATA CAGCTATTTT TAATGTTGTG	7020
TTTGAGCCTA ATAGTTGTAA TTTACAAATA AAAAGAGAAT ATTTTACGAG TTTTAATAAG	7080
AATATTTTAT GTTGCCGTAT AGTGTCATCA GTTCAAAACA CATTAAATTT AAACATTAAT	7140
TTGGGTAGAA ATAAACGGTT TAATGACGAA GTATCTAAAC TGGATTCAAG TACAATTTTA	7200
ATGTCGGCCT CTGCTGGAGG TAGAAAAGGT GTTCAGTTTA AAGTAGTATG TCATTCTAAG	7260
GTTACGGATG GTGAAGTAAG TGTATTGGGA GAGACAATAG TTATTCGGAA TGCTACAGAG	7320

547

GTATTTCTTT ATCTCAAATC AATGACGGAT TATTGGGGAA ATATAGATAT TTCTTCTCTT	7380
CAGGGAGAAT TTAGTAGTAT TGATTACTTT ACAGAAAAAG ATGAACATGT AAAAAAATAT	7440
CAGGAGCAAT TTAATAGAGT TGATTTTAAA CTAGACTATA GTAAAGGTTG TCTTAGCATT	7500
CCAACGAATC TACTTCTTGA AAACACTAAA AAGTATAGTA ACTACTTGAC TAACTTGTTA	7560
TTTCATTATG GAAGATATCT GTTAATATCG TCTAGTCAAC CGAATGGTTT ACCTGCCAAT	7620
CTTCAAGGAA TATGGTGTGA TGAATTAAAT CCAATTTGGG GTTCTAAATA TACGATTAAT	7680
ATTAATACTC AAATGAATTA TTGGATGGTA GGTCCATGTG ATTTACCAGA AGTAGAATAT	7740
CCATTATTTG ATATGCTCGA AAGAATGAGA GAACCGGGAA GACTAACCGC TAAGAAAATG	7800
TATGGAGCTA GAGGTTTTAC AGCACATCAT AATACGGATG GTTTTGGCGA TACGGCTCCC	7860
CAATCTCATG CCATGGGGGC TGCAATTTGG GTATTAACTA TTCCATGGTT ATGTACTCAT	7920
ATTTGGGAAC ACTATTTATA TTCCAAGAT GAGCGTATTC TTACGGAACA TTTTGAAATG	7980
ATAAAGAAG CATTTCTTTT CTTTGAAGAT TATTTATTTG AGGTGGATGG CTAATTGATG	8040
ACAGGTCCAA GTGTCTCACC GGAAAATAAA TATCGCTTAA AAAATGGTAT TGAAGGAAAT	8100
GCTTGTCTAT CATCTACAAT TGATAATCAA ATTCTAAGAT ATTTTGTGA TTCATGCATT	8160
GGCATTGCAA AACAAATTAGG AGACAATTCG GATTTTATTA GTCGTGTGAA GGAGTTAAAA	8220
AAGAACTAC CTAAAACAAA AATAGGTAGT AATGGGCAAA TCCAAGAATG GTTAGAAGAT	8280
TATGAAGAAG TAGAGCCTGG GCATAGACAC ATTTCCACCTC TATTTGGGCT TTATCCTTAT	8340
AATGAGATTG ATATTCATAA AACTCCGGAA TTAGCAGAAG CAGCTAAAAT CACTATCAAT	8400
AGGAGATTAT CAAACGCTAA TTTTATCTC TCACAGGAGA GGGAGCAAGC GATTAATAAT	8460
TGGTTAGTAA GTGGTTTGCA TGCTAGTACA CAAACAGGTT GGAGTGCTGC ATGGCTGATT	8520
CATTTTTTTG CGAGACTATA TCAAGGTGAA CCTGCTTATA ACCAGATTAA TGGTTTGTTA	8580
AATAATGCCA CTCTTGCCAA TTTATTTCTT GACCATCCAC CATTTCAAAT TGATGGTAAT	8640
TTAGGTTTGG TGAGTGGAAT TTGTGAATTA TTAGTACAGA GCCATCATAA TTGGTTATCA	8700
CTAATTCAG CTTTACCTTC TGCTTGCTCA GAAGGAGAAG TGAAAGGTTT CAGAGTAAGA	8760
GGAGGATATA AGGTATCGTT TGCTTGGAAG AATGGGGATA TAACATTCCT AAAATTGGAA	8820
GGAGGAAACA AAGATCAAAA AGTAAGAGTA AGAATATATG GCAAAAATAC TGATGTACAA	8880
AATATTGAAT TGGTATTTAA TTCAGAAAAA ATTATTGAGT TAAATTTTAA GGTATAAGTC	8940
ATGAATAAAG AAAAAATAAA AAGAAAATTA ATCACAATAT TGTGTGTATG TATTGGGATG	9000
TTATGTTTTG GATTGTTAGC AGGAGTTAAG GCTGATAATC GTGTTCAAAT GAGAACGACG	9060

648

ATTAATAATG AATCGCCATT GTTGCTTTCT CCGTTGTATG GCAATGATAA TGGTAACGGA	9120
TTATGGTGGG GGAACACATT GAAGGGAGCA TGGGAAGCTA TTCCTGAAGA TGTAAAGCCA	9180
TATGCAGCGA TTGAACTTCA TCCTGCAAAA GTCTGTAAAC CAACAAGTTG TATTCCACGA	9240
GATACGAAAG AATGAGAGA ATGGTATGTC AAGATGTTGG AGGAAGCTCA AAGTCTAAAC	9300
ATTCCAGTTT TCTTGGTTAT TATGTCGGCT GGAGAGCGTA ATACAGTTCC TCCAGAGTGG	9360
TTAGATGAAC AATTCCAAAA GTATAGTGTG TTAAGAGGTG TTTTAAATAT TGAGAATTAT	9420
TGGATTTACA ATAACCAGTT AGCTCCGCAT AGTGCTAAAT ATTTGGAAGT TTGTGCCAAA	9480
TATGGAGCGC ATTTTATCTG GCATGATCAT GAAAAATGGT TCTGGGAAAC TATTATGAAT	9540
GATCCGACAT TCTTTGAAGC GAGTCAAAAA TATCATAAAA ATTTGGTGTG GGCAACTAAA	9600
AATACGCCAA TAAGAGATGA TGCGGGTACA GATTCTATCG TTAGTGGATT TTGGTTGAGT	9660
GGCTTATGTG ATAACTGGGG CTCATCAACA GATACATGGA AATGGTGGGA AAAACATTAT	9720
ACAAACACAT TTGAACTGG AAGAGCTAGG GATATGAGAT CCTATGCATC GGAACCAGAA	9780
TCAATGATTG CTATGGAAT GATGAATGTA TATACTGGGG GAGGCACACT TTATAATTC	9840
GAATGTGCCG CGTATACAT TATGACAAAT GATGTACCAA CTCCAGCATT TACTAAAGGT	9900
ATTATTCCTT TCTTTAGACA TGCTATACAA AATCCAGCTC CAACTAAGGA AGAAGTTGTA	9960
AATAGAACAA AAGCTGTATT TTGGAATGGA GAAGGTAGGA TTAGTTTATT AAACGGATTT	10020
TATCAAGGAC TTTATTGAA TGATGAAACA ATGCCTTTAT ATAATAATGG GAGATATCAT	10080
ATTCTTCCTG TAATACATGA GAAAATTGAT AAGGAAAAGA TTTCATCTAT ATTCCCTAAT	10140
GCAAAAATTT TGAATAAAAA TAGTGAGGAA TTGTCTAGTA AAGTCAACTA TTTAAACTCG	10200
CTTTATCCAA AACTTTATGA AGGAGATGGG TATGCTCAGC GTGTAGGTAA TTCCTGGTAT	10260
ATTTATAATA GTAATGCTAA TATCAATAAA AATCAGCAAG TAATGTTGCC TATGTATACT	10320
AATAATACAA AGTCGTTATC GTTAGATTTG ACGCCACATA CTTACGCTGT TGTAAAGAA	10380
AATCCAAATA ATTTACATAT TTTATTGAAT AATTACAGGA CAGATAAGAC AGCTATGTGG	10440
GCATTATCAG GAAATTTTGA TGCATCAAAA AGTTGGAAGA AAGAAGAATT AGAGTTAGCG	10500
AACTGGATAA GCAAAAAATTA TTCCATCAAT CCTGTAGATA ATGACTTTAG GACAACAACA	10560
CTTACATTAA AAGGGCATAC TGGTCATAAA CCTCAGATAA ATATAAGTGG CGATAAAAAAT	10620
CATTATACTT ATACAGAAAA TTGGGATGAG AATACCCATG TTTATACCAT TACGGTTAAT	10680
CATAATGGAA TGGTAGAGAT GTCTATAAAT ACTGAGGGGA CAGGTCCAGT CTCTTTCCCA	10740
ACACCAGATA AATTTAATGA TGGAATTTG AATATAGCAT ATGCAAAACC AACAACACAA	10800
AGTTCTGTAG ATTACAATGG AGACCCTAAT AGAGCTGTGG ATGGTAACAG AAATGGTAAT	10860



649

TTTAACTCTG GTTCGGTAAC ACACACTAGG GCAGATAATC CCTCTTGGTG GGAAGTCGAT	10920
TTGAAAAAAA TGGATAAAGT TGGGCTTGTT AAAATTTATA ATCGCACAGA TGCTGAGACT	10980
CAACGTCTAT CTAATTTTGA TGTGATTCTA TATGACAATA ATAGAAACGA AGTTGCTAAG	11040
AAACATGTTA ATAATTTGTC GGGTGAATCT GTTAGTCTAG ATTTCAAAGA AAAAGGAGCA	11100
AGGTATATTA AAGTTAAATT ACTAACGAGT GGAGTGCCTT TGAGTTTAGC AGAAGTAGAG	11160
GTTTTTAGAG AATCAGATGG TAAGCAATCT GAAGAGGATA TAGATAAAAT AACAGAAGAT	11220
AAAGTAGTCT CTACAAATAA GGTAGCTACT CAAAGTTCAA CCAATTATGA GGGTGTAGCT	11280
GCTTTAGCAG TTGATGGTAA TAAAGATGGA GATTACGGAC ATCATTCCGT GACTCATACT	11340
AAGGCAGATT CTAACGCTTG GTGGCAGGTC GATCTGGGAG AAGAGTTTAC GGTTTCTAAA	11400
GTTGATATTT ATAATAGAAC AGATGCCGAA CCTCAGCGTT TATCTAATTT TGATGTTATT	11460
TTTCTATCTT CATCAGGAGA AGAAGTTTTT AGAAGACATT TTGATAAAGT AGTTGATGGT	11520
TTGTTATCTT TAAAAGTACC TTCTGTAGGG GCTAAGCTAG TCAAAATAGA ATTAAATCA	11580
GCAGCTATTC CGTTAAGTTT AGCGGAAGTT GAAGTCTATG GTTCAAAGAG AACTCCGAAC	11640
AAACTTTCTA ATATTCGATT AACAAAAGAA ACTCGACAGA GTTCAACCGA TTACAATGGT	11700
TTTCTCTGTC TAGCAGTTGA TGGAAATAAA AACGGAGATT ATGGTCATCA TTCAGTGACT	11760
CATACCAAAG AAGATTCTCC TTCATGGTGG GAGATAGATT TAGCACAAAC CGAAGAATTA	11820
GAAAAGTTAA TTATTTATAA TAGAACAGAT GCTGAAATTC AGAGATTATC AAATTTTGAT	11880
ATTATTATAT ATGATTCAAA TGATTATGAA GTTTTACAC AACATATTGA CAGTTTAGAA	11940
AGCAATAATC TATCCATAGA CTTAAAAGGA CTGAAGGGAA AAAAGGTTAG AATTCTTTTC	12000
AGAAGCGCAG GAATTCCTTT AAGTTTAGCA GAGGTAGAGG TTTATACTTA TAAGTAATTT	12060
TAAAAATTAT CACCCAGGCT ACCGTAAATA TAATGGAGAT GGTAGTATGA AAGAAACAGA	12120
AAAAAAGAG GAAAATAGTA TGATTCAACA TCCACGTATT GGGATTCTGC CGACTATTGA	12180
TGGTCGTCGT CAAGGTGTAC GCCAATCACT TGAAGTGCAA ACAATGAACA TGGCTAAAAG	12240
TGTGGCAGAT TTGATTTCAA GCACATTGAA ATATCCAGAT GGGGAACCTG TGGAATGCGT	12300
GATTTCTCCA TCTACTATTG GCCGTGTACC AGAGGCTGCA GCTTCCCATG AGTTGTTTAA	12360
AAAATCAAAT GTTTGCGCAA CAATTACAGT TACACCATGC TGGTGTTATG GTAGTCAAAC	12420
TATGGATATG TCTCCAGATA TTCCTCATGC TATTTGGGGA TTAAATGGGA CAGAACGCCC	12480
AGGAGCTGTC TATCTTGACG CTGTACTAGC TTCACATGCT CAAAAAGGGA TTCCAGCCTT	12540
TGGGATTTAT GGAAGAGATG TTCAGGAAGC TAGTGACACA GATATTCAG AAGATGTCAA	12600

650

AGAAAACTT TTACGCTATG CGCGTGCAGC TCTTGCAACT GGCTTGATGA GAGACACTGC	12660
TTACCTATCA ATGGGTAGTG TTTCGATGGG GATTGGTGGT TCTATTGTAA ATCCGGATTT	12720
CTTCCAAGAA TACTTAGGAA TGCGAAATGA ATCGGTAGAT ATGACGGAGT TCACGCGCCG	12780
TATGGACCGT GGTATTTACG ACCCTGAAGA GTTCGAACGT GCGCTCAAAT GGGTGAAAGA	12840
AAACGTAAAA GAAGGATTCG ACCATAACCG TGAAGACCTT GTTTTAAGCC GTGAAGAAAA	12900
AGATAGACAA TGGGAATTTG TTATTAAGAT GTTCATGATT GGACGTGACT TAATGGTTGG	12960
TAACCCAAGA CTTGCTGAAC TTGGTTTGA GGAAGAAGCG GTTGGTCACC ATGCTTTAGT	13020
AGCTGGTTTC CAAGGTCAAC GTCAGTGGAC AGACCATTTT CCAATGGGG ACTTTATGGA	13080
AACTTTCCTC AATACTCAGT TTGACTGGAA TGGTATTCGA AAACCATTTG TATTTGCGAC	13140
AGAGAATGAT TCACTAAATG GTGTGTCTAT GCTCTTTAAT TATCTATTAA CAAATACTCC	13200
ACAAATCTTT GCTGATGTGC GTACTTATTG GAGCCCAGAG GCTGTTAAAC GTGTAACGGG	13260
ACATACTTTA GAGGGTCGTG CTGCAGCTGG CTTCTTACAT CTAATCAACT CTGGTTCTTG	13320
TACATTGGAT GGTACAGGTC AAGCTACTCG AGATGGCAAA CCTATTATGA AACCATTCTG	13380
GGAGTTGGAA GAAAGTGAAG TGCAGGCTAT GCTTGAAAAT ACAGACTTCC CACCAGCAAA	13440
CCCGGAATAC TTCCGTGGAG GAGGATTCTC AACTCGTTTC TTGACGAAGG GGGATATGCC	13500
AGTAACAATG GTACGTCTCA ATCTTCTAAA AGGGGTTGGT CCAGTGCTAC AAATTGCAGA	13560
AGGTTACACA CTGAACTTC CTGAAGATGT TCACCATACT TTAGATAATC GTACAGATCC	13620
AGGATGGCCA ACTACTTGGT TTGCTCCAGC TTGACAGGA AAAGGTGCTT TCAAGTCTGT	13680
CTATGACGTC ATGAATAATT GGGGAGCTAA TCACGGAGCC ATAACATATG GACACATTGG	13740
AGCAGACTTG ATTACCTTGG CTTCTATGTT GAGAATTCCT GTCAATATGC ATAATGTACC	13800
TGAGGAAGAT ATCTTTAGAC CTAAAAATTG GTCCTTATTT GGAACAGAAG ATCTAGAATC	13860
AGCAGACTAT CGTGCATGTC AGTTGTTGGG GCCACTACAT AAATAAACT TGTTTATATA	13920
GGAGGTGAAC TTACGTCCCT CCTATCCTTT TAAAAAGATT TGTAAACAA TTCACAAA1A	13980
ATTGAAAACG AATACAAAAA GTAATATAAT GATGTTAAAT AGATAGCGCG GAGGCGCAGG	14040
AGGAAAATTA TATGGCTATA TTTTATGTTT CGGCAGTCAA CCTTATTGGA AAAGGTGTTG	14100
TAAATGAAGT GGGTCCTTAT ATCAAGGAAC TTGGCTATAA AAAGGCACTT TTGGTGACAG	14160
ATAAGTACAT CGAAGGCAGT GATATTTTAC CTAAGACTTT AAAACCACTG GATACAGAAG	14220
GAATCGAATA T	14231

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:

651

(A) LENGTH: 16995 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGTTCTCTTA ACTTTTTTAG GATGGCATT CCGCTCTCA GGTACTCATT TTCTGCTgAA	60
GACGTTCTAA TTCTGTCCTC TCTTCAGGTC TCGTTTTTGG CTTACGTCCC ATTTTAGGTA	120
CTCTCCCTCT TGTMTTCTCA ACAATAGTAT ACCCGTTTTT CCTGTATTGT GCTAGCCAGT	180
TAAGAAGTAT CGTACGACTT GGGAGACCGT ATTCAAGAGA AACTCTATCT TTAGTCCAGC	240
CTTCATGTCA GACTTTATTA CTCATTTCTT GTTTTAAATC AGGAGAATAG TAACGATTTT	300
TTCTTTTTTT GACGAACTCT ATTCCGTAAC GATCAATCAA TTTAATCATG TACCTAATAT	360
TAGAATTGCT TATCCCAAAT TTATTTGAAA GCTTCTCTAA GCTATATCCT TGTTTTCTAA	420
GTTCATAGAT CTGAACCTTA TCATCATAAG TTAGTTTCAT AATAAAAACA CCCCCAAAGT	480
TAGATTTTTT CTGCTAACT TTTGGGGTGT AGTTCATGTA CACCTGATAT GATGCGTTTT	540
ATAATTTTTA AGCCTTTTTG CCCAGCCTCG TCAAAAGTAA TCTTTTGACA CAAAATCTGT	600
GACAAAACCT TAGTTTAAAA GGTTTTAAAC TTTGTATATA CTAGTTTTAA GAAAAGGAGG	660
ATGATCTAAT GGAAGAAAAA GTATCATTGA AAGTCAGGGT TCAAAAAC TA GGGACATCGC	720
TTTCAAATAT GGTATGCCC AATATTGGAG CATTATTGCG TTGGGGAGTA TTGACTGCCC	780
TCTTTATCGC TGATGGCTAT CTGCCAAATG AACAGTTAGC TACTGTTGTT GGTCTATGT	840
TAACGTATTT ATTGCCAATC CTGATTGGTT ACACAGGTGG ATATATGATC CATGGCCAAC	900
GTGGTGCCGT TGTAGGAGCT ATTGCTACTG TTGGTGCAAT CACAGGTTCT AGTGTTCTTA	960
TGTTTATCGG AGCTATGGTA ATGGGCCCAC TGGGAGGATG GACTATCAAG AAATTTGATG	1020
AGAAGTTCCA GGAATAAATT CGTCCCGGAT TTGAAATGTT AGTTAATAAC TTCTCAGCTG	1080
GTCTCGTTGG TTTTGCAATTA TTGCTTTTGG CTTTCTACGC AATCGGTCCA GTCGTATCGA	1140
CTCTTACTGG AGCTGTTGGG AATGGTGTG AGGCTATTGT CAATGCTCGC CTCCTTCCTA	1200
TGGCTAATAT TATCATCGAA CCGGCTAAAG TCCTTTTCCT CAATAATGCC CTCAATCATG	1260
GCATTTTAC TCCTCTGGGA GTAGAACAGG TAGCTCAAGC TGGTAAGTCA ATTCTCTTCC	1320
TATTGGAAGC TAATCCTGGA CCAGGTCTGG GAATCTATT ACCTTATGCT GTATTCGGTA	1380
AAGGTTCTGC TAAATCTTCT TCTTGGGGGG CAATGGTTAT TCATTTCTTC GGAGGGATTC	1440
ATGAAATTTA CTTTCCTTAT GTTATGATGA AGCCTACTCT ATTTTtagCT GCTATGGCAG	1500

652

GAGGTATCTC	TGGAACCTTT	ACTTTTCAAC	TCTTAGACGC	TGGTCTTAAA	TCTCCAGCTT	1560
CACCAGGTTT	TATTATTGCG	ATTATAGCTA	CGGCGCCAAA	AGGTGTTTGG	CCCCATCTAA	1620
ATGTTCTTTT	AGGTGTTTTA	GTGGCAGCAG	TTGTTTCTTT	CCTTGTAGCA	GCCCTTATTC	1680
TTTCATGCAGA	CAAGTCAACT	GAGGATTTCG	TCGAAGCTGC	TCAGGCGGCT	ACCCAAGCAG	1740
CTAAGGCTCA	GTCTAAAGGT	CAGTTAGTAT	CAACTTCTGT	TGATGCAGTT	GTTTCGACAG	1800
ACTCAGTGGA	AAAAATCATT	TTGCGCTGCG	ATGCTGGTAT	GGGAAGCTCT	GCTATGGGAG	1860
CTAGTATTCT	TCGAGATAAG	GTTAAAAAAG	CAGGTCTAGA	GATTCCAGTA	TCTAATCAGG	1920
CAATCTCAAA	TTTGCTTGAT	ACACCAAAAA	CATTAAATTGT	TACTCAGGAA	GAAGTGACAC	1980
CAAGAGCTAA	AGACAAGAGT	CCAAGTGCTA	TTTCATGTTT	TGTTGATAAT	TTCTTAGCGT	2040
CCTCTCGTTA	TGATGAAATT	GTAGCTTCAT	TAACAGGAGC	TTCTCCAATA	GCAGAAATTG	2100
AAGGAGATAT	ACCAACTTCA	GCACCAGTAG	ATAGTCAGGA	AAGTGACCTT	AACCATATTG	2160
ATGCTGTAGT	AGTTCCTTAT	GGTAAAGCAC	AGGGAAGTGC	AACTATGGGC	TGTGAAACGA	2220
TTGCGGCTAT	TTTGTAGAAC	AAGAATATTC	GTATTCCAGT	TTCTACTGCC	AAAATTTTCT	2280
AAATAGGTGA	ATTTAATTCT	AAAAACATAA	TGATTGTAAC	AACTATTTCT	TTACAGGCAG	2340
AAGTGACGCA	AGCAGCACCG	AATTCTCAAT	TTCTTATTGT	GGATAGTTTA	GTAACAACAC	2400
CAGAATATGA	CAAAATGGCT	GCTAGAATGT	ACAAATAGAA	CTAGAGGTTT	CTAAATTACG	2460
AATGCTATTA	ACCAAACGAG	AAGAACAATT	ATTGAAGGCT	TTCTTACATG	TAGGGAAGCT	2520
TTCAATGCAA	GATATGACTG	AAATCTTACA	GGTTTCATCT	AGAACAATTT	ATCGAACTTT	2580
ATCAGATTTG	ACAGATAGCA	TGGAGCAATA	TGGAATCGAA	ATAACGAAGC	ATGGGAAATA	2640
CTATATTTTG	ACTGGAGAGT	TGGATGATTT	GCCGACAGAA	CTTGAAGTGT	TAGTTGAGTA	2700
TAGTCCCCAA	GAAAGACAAG	AGTTGATTAC	CTATCGCCTT	CTGACTGAGA	GTGGTTTTGT	2760
CACCAATGAA	GCATTGCAAG	AGTGACAGAA	AGTCAGTAAT	GTAAGTATTA	TTTCAGGATAT	2820
TTTCAGATATT	GATAAGCGTC	TTTTAGACTT	TGATCTGAAA	ATTGAACGAC	AAAAAGGTTA	2880
TCGGATTTCT	GGTGATTTCAG	TTGGTAAGAG	AAGATTTTTG	GCTATTTTAC	TGACAAACTG	2940
TATCTCAGTA	GCAGATTTTT	CAACCGGTAA	TTTTGGGAGC	TTTGATATTT	TAGAAGCAGA	3000
TAGAAGTGGG	CTGGCCAGTC	AGATTGTTAA	TAAGCAACTG	TCAGGTTTTT	CAGATATGGA	3060
TGCTAGGATG	AAGATGTTTT	TTGCGATCTT	GTTATCTCTT	ATAGGTCAGG	AGCAAAACAT	3120
TGAAAATTCA	CCTAATACTA	GTAAGCAGGC	TTTGGAAATT	TCTCAAAAAA	TTTTTCAAGC	3180
TTACTCTAAG	CAGACTGCAC	AATTTTATAG	TATTCAGGAA	ATTATCTATT	TTGCGAGCAT	3240
CTTGATGAA	TTAATCATT	AACGTCAGGA	CAATCCGCTC	TTTACGGAGA	AATTTGATGG	3300

653

TGAATTTTTC TACAATATTT CAAATCTGAT TGATACGGTT TCCATGTATA CCAAGATTGA	3360
CTTTTTTAAG GACAAGGTTT TATTCAATTT TCTTTTCCAT CATATTCGGC TCAGTTTAGG	3420
CGTCCCTATC CTTTTTCAGG GTGAAAATTT GCCAGAATCT ATCCAGATTT TAGTTGAAAG	3480
GAATAAATTT CTTTATACAG TCATCAGTCT TTTAGTGAAT GATATTTTTC CGAAATATCT	3540
TCATACAGAG TATGAGTATG GCATGATTGC CCTACATTTT ATCTCTAGCT TAGGCCGTAG	3600
TCCAGAGATT TATCCAGTCC GTGTTTGTCT TTAAACGGAT GAACGTCGGG TCACTAGAGA	3660
TTTATTAGTC AGTAAATTA AGAGTGTTC TCCTTTTGTA GAGTTGATAG ATATTCAATC	3720
TCTAGTAGAT TACCACAGTA TTGATCTCAG TCAGTATGAT TATATTTTAT CTACCAAGCC	3780
GCTGACTAAT CAGGAAATCG ATGTAATTC TAGTTTTCCA ACCGTCAAAG AATTGCTTGA	3840
ATTACAGGAA CGACTTCAGT ATGTACAGGC ACATCGTACA ATTGTCGGCG GTGATGCTAT	3900
CGCTCCAGAG AAAAGTTATG ACTTGCAAGA TTATTTAATA TCTAGTAGTC AGCTTTTGAG	3960
TCAATTCGAG TTGGTTCAAT TGGAGAATAA TCAATCATTT GAGCACACGG TAGAACAAAT	4020
CATCCAATAT CAGAAGAATG TGAGTGACAG AGCTTACCTA ACAAGAAAAT TGTTATCTCA	4080
CTTCCAGAAT AGTCCTATGG CTATTCCTAA TACTGGTCTG GTGCTTTTAC ATAGTCAGTC	4140
TAGCAAAGTA ACAACAAATA GTTTTACTAT GTTTGAACTC AACTACCTA TCTCCGCATT	4200
GTCAATGAAA CGAGAGGAAG AAGAGGTCAA AAGGTGTCTG CTAATGCTAA TGTCTAAAGA	4260
AGCTAGCGAG GAAGCTAGAG ATTTAATGAC AGCTATTAGT CAGTCGATTA TTGAAAATCA	4320
TCTTTATACA GAGATTTACA AGACGGGAAA TCAATCCATT ATTTATCAGA TGCTAAATAC	4380
TATTTTTAAC GAAAAAATTA AGAAATTGGA GAACTAATAT GAAACTTGAA AAACATTTGA	4440
TTAAGCTTAA TAAACAATTT TCTAACAAGG AGGAAGCTAT TTGTTATTGT GGGCAAGTTC	4500
TTTATGAGGG TGGATATGTT AATGAAGACT ATATTGAAGC CATGATTGAG CGAGATAAAG	4560
AGCTATCTGT TTACATGGGT AACTTTATCG CCATACCGCA TGGAACAGAT GCAGCAAAAA	4620
ATGATGTCCT CAAGTCTGGT ATTACAGTCG TTCAAGTCCC TAGAGGGGTT GATTTTGGA	4680
ATGTATCTAA CCCTCAAGTG GCAACGGTTC TTTTGTGAT TGCTGGTATT GGTAATGAAC	4740
ACTTAGAAAT TATTCAGAAA ATTTCTATCT TCTGTGCAGA TGAGATAAT GTTCTTAAAC	4800
TAGCAGATGC TCACTCAAAA GAGGAAGTAT TGCGCTTATT TGATGCTGTT GAATAATTGA	4860
ATTTAGTCAT TTGTCATCTA GTATATATGT CCCTCAAATA GGAAAAGGAG AAATTGAATG	4920
AAACATTCTG TTCATTTTGG TGCCGGTAAT ATCGGTCGTG GTTTTATAGG TGAAATCTTA	4980
TTTAAAAATG GTTCCATAT TGATTTTGTG GATGTCAATA ATCAGATAAT TCATGCTCTG	5040

654

AATGAAAAGG GCAAGTATGA AATTGAAATT GCACAGAAAG GACAGTCTCG TATAGAAGTA	5100
ACTAATGTGG CTGGCATTAA TAGCAAAGAA CATCCTGAGC AAGTCATTGA AGCGATTCAA	5160
AAGACGGATA TTATTACTAC TGCAATCGGA CCTAATATAC TCCCTTTTAT CGCCGAAGTT	5220
CTAGCCAAAG GAATCGAAGC TCGCCGAGTT GCAGGAAATA CACAGGCATT GGATGTTATG	5280
GCCTGTGAAA ATATGATTGG CGGGTCTCAA TTTCTTTATC AAGAAGTCAA GAAATATTTA	5340
AGTCCGGAAG GTTTGACATT TGCTGATAAC TACATAGGTT TTCCAAATGC TGCAGTAGAC	5400
AGGATTGTTT CAGCACAAAG TCACGAAGAT TCCCTTTTGT TTGTGGTCGA GCCCTTTAAT	5460
GAATGGGTCG TGGAAACCAA GCGTCTTAAA AATCCAGATT TACGTCTAAA AGATGTGCAT	5520
TATGAAGAAG ATTTAGAACC CTTTATTGAG CGAAAACTTT TTTCAGTCAA TTCTGGACAT	5580
GCAACTTCAG CTTACATTGG TGCGCATTAT GGTGCCAAGA CAATTTTGGG AGCTCTTCAA	5640
AATCCTAATA TTAAATCTCG GATTGAATCT GTATTAGCTG AAATTCGGAG TCTCTTGATT	5700
GCCAAATGGA ACTTTGATAA AAAAGAATTG GAGAATTATC ACAAAGTCAT TATAGAACGA	5760
CTTGAAAACC CTTTCATAGT GGACGAGGTT AGTCGCGTAG CTCGTACTCC AATCCGAAAA	5820
TTAGGCTATA ATGAACGATT CATCCGGCCG ATACGTGAAT TGAAAGAACT CAGTTTGTCA	5880
TATAAAAACC TACTTAAAC AGTTGGCTAT GTCTTTGACT ATCGCGATGT AAATGATGAA	5940
GAAAGTATTC GATTAGGTGA ATTGTTGGCT AAACAATCAG TCAAAGATGT TGTATACAA	6000
GTTACAGGTT TAGACGACCA AGAATTGATT GAGCAAATTG TAGAGTATAT TTAATCTTTT	6060
TCGAAAATCT CTTCAAATCA GGTTAGCATC GCTTTGTCTT AGGCATATGT TGTCTATCT	6120
ACAACCTCAA AGCAGTGCTT TGAGCTGACT CCGTCAGTCT TATCTGCAAT CTCAAAACAC	6180
TGTTTGAGTT ATCTGCGGTA ATCTTTCTAG CTTGTCTTTG ATTTTGTGTT TTATTTATAA	6240
GGTAAAAGAA GCTGGACAAA AAGTCTTCAA AATCGGGAAA AGGCAGCCTA TCGGGTGTTC	6300
AAAAATCTTG ATAGGATGTC CTTTATTATG GAAAGCCTTA TTGGATTTTC TCCTCAGATT	6360
GAGTTTTTGA TCAGCTTTAT GAGATAGGTC TTGCTAGAGA TGTAGCCCAT CATGTTATTT	6420
TTATGGACAG TGGGAAAATT GTTGAAAAAA ATAATGCCCA TCAATTCTTT AGTCGTCCAA	6480
GAGAAGAACG AACCAAGCAA TTTTGGAACG AATTCCTTCG AATGCGATCT ATATAGTAAA	6540
ATGAAACAAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTTCTAA AAATGTTTTA	6600
GAAGTAGAGG TGTACTATTC TAGTTTCAAT CTACTATATA ACTGAAAAAT TAGATAAATT	6660
AGTTTTGGAA AATGACTAAC CAAAAGATAT CCAAAGTAGT CTAATAATTGT CTATACTTTA	6720
TGACTGTTTT AGTTAGGAAA AAGGCTTGTT GTCTATAATT GTCTGCATTA GTCTAGATTT	6780
TATTTATAGA AAATGTTATA ATAGACTGTA TTTAAAAAAT TTTAAGGAGA AATGACAGAA	6840

655

TGTCTGTATC ATTTGAAAAC AAAGAAACAA ACCGTGGTGT CTTGACTTTC ACTATCTCTC	6900
AAGACCAAAT CAAACCAGAA TTGGACCGTG TCTTCAAGTC AGTGAAGAAA TCTCTTAATG	6960
TTCCAGGTTT CCGTAAAGGT CACCTTCCAC GCCCTATCTT CGACCAAAAA TTTGGTGAAG	7020
AAGCTCTTTA TCAAGATGCA ATGAACGCAC TTTTGCCAAA CGCTTATGAA GCAGCTGTAA	7080
AAGAAGCTGG TCTTGAAGTG GTTGCCCAAC CAAAAATTGA CGTAACTTCA ATGGAAAAAG	7140
GTCAAGACTG GGTATCACT GCTGAAGTCG TTACAAAACC TGAAGTAAAA TTGGGTGACT	7200
ACAAAAACCT TGAAGTATCA GTTGATGTAG AAAAAGAAGT AACTGACGCT GATGTCGAAG	7260
AGCGTATCGA ACGCGAACGC AACAACTGG CTGAATTGGT TATCAAGGAA GCTGCTGCTG	7320
AAAACGGCGA CACTGTTGTG ATCGACTTCG TTGGTTCTAT CGACGGTGTT GAATTTGACG	7380
GTGGAAAAGG TGAAACTTC TCACTTGGAC TTGGTTCAGG TCAATTCATC CCTGGTTTCG	7440
AAGACCAATT GGTAGGTCAC TCAGCTGGCG AAACCGTTGA TGTTATCGTA ACATTCCCAG	7500
AAGACTACCA AGCAGAAGAC CTTGCAGGTA AAGAAGCTAA ATTCGTGACA ACTATCCACG	7560
AAGTAAAAGC TAAAGAAGTT CCGGCTCTTG ACGATGAACT TGCAAAAGAC ATTGATGAAG	7620
AAGTTGAAAC ACTTGCTGAC TTGAAAGAAA AATACAGCAA AGAATTGGCT GCTGCTAAAG	7680
AAGAAGCTTA CAAAGATGCA GTTGAAGGTG CAGCAATTGA TACAGCTGTA GAAAATGCTG	7740
AAATCGTAGA ACTTCCAGAA GAAATGATCC ATGAAGAAGT TCACCGTTCA GTAAATGAAT	7800
TCCTTGGGAA TTTGCAACGT CAAGGGATCA ACCCTGACAT GACTTTCCAA ATCACTGGAA	7860
CTACTCAAGA AGACCTTCAC AACCAATACC AAGCAGAAGC TGAGTCACGT ACTAAGACTA	7920
ACCTTGTTAT CGAAGCAGTT GCCAAAGCTG AAGGATTTGA TGCTTCAGAA GAAGAAATCC	7980
AAAAAGAAGT TGAGCAATTG GCAGCAGACT ACAACATGGA AGTTGCACAA GTTCAAAACT	8040
TGCTTTCAGC TGACATGTTG AAACATGATA TCACTATCAA AAAAGCTGTT GAATTGATCA	8100
CAAGCACAGC AACAGTAAAA TAATCTTAAT AAACAGAAAA CCCACCTGAA TTGGTGGGTT	8160
TTCTGATGCA CTATTTTCCA AAAATCTCTT TGAGGTCTGT GTCTGTAATC CCAATCATGG	8220
CTGGGATGCG GTCCCAGTTT TCTTCGGTTA GGATGTAGGA TTGTTTCAGAG GCACTTGATG	8280
TGACTGTTTC AGAGACAGCT TGTGCTTTT CTTCAACATT CTCCAGTAGA TCACTGAAGC	8340
GTTCAATCAG ATAGGTTTTT CGGGCAGTTC CGATGTGTTG GGTAGCATAG TCGAAGGCTT	8400
GTAATTCGCC TAGTAAGATG AGTTTGCTTT TGGCACGTGT AATGGCTGTG TAGATGAGAT	8460
TTCGCTCCAG CATACGTCGG CTAGCACTAG TAATCGGTAG GATGACAACT GGGAACTCAC	8520
TTCCCTGAGA CTTATGAATA CTCATGGCAT AGGCCAAGCG AATCTTGATC CATTGCTTAC	8580

656

GGGGTAAGA GACTTCATTA CCATCAAAAT CAATGACAAT CTCGTCTTGT TTCGATTCCG	8640
TGTATTTACC AGGAATCAGG TCTGTGATAG CTCCTAAATC CCCATTAAAG ACATTGATTT	8700
CAGCATCGTT AACCAAATGA ATGACCCTGT CTCTCTTACG ATAGTGACAC TGAGGAGCTT	8760
CAAAACTGAG TTGATCTTTT TGTGGGGGAT TGAGCAGGTC TTGCATGAGC TGATTGATAG	8820
CATCAATCCC TGCCGTCCCT CGGTACATAG GAGCCAGAAC TTGGATATCA CGGGCGGGAA	8880
TACCATTTCT GAGGGCGGCA CCTAAGATTT TTTCAATGGT GGCAGGAATA TGGCCACTAG	8940
CAATTTCAAA GTAGGAACGG TCAGCTTTTT TTTGGGTGAA ATCAGCTGGC AAGATGCCCT	9000
GTCGAATCTG ACTAGCTAGG GTGACGATGG TTGATTCTTT GCTTTGTCTG TAAATTTTTT	9060
CCAAGCGAGT CTGAGGAATC AAAGGAATAT GAAGTAGATC CGCTAGAACC TGTCCAGGAC	9120
TGACAGAAGG TAGCTGATCA CTGTACCTA CGATGAGGAT CTTACTGTTA GAAGAGATAT	9180
TGGAGAAGAG TTGATTGGCC AGCCAAGTAT CTACCATAGA GAATTCATCC ACGATGATAA	9240
AGTCAGCATC TAGGTAATCT TCCAGATGAC TGGTATCATC GTCACCTGTC ATTCCCAAGT	9300
GGCGATGTAT GGTGCGGCTA GGCAAACCTG TCAATTCATT CATGCGACGA GCAGCTCGAC	9360
CAGTTGGAGC AGCAAGAAGA ATGGGCAGAT TGCTTTTCTT CCTGAAGTCA AGTCCTTCTA	9420
AAAGGGCATA AACAGCAATG ATTCCATTGA TAACAGTTGT CTTACCAGTA CCAGGCCCCAC	9480
CTGTCAAGAT AAAGACCTTA TTCTGGATAG CATCACAGAT AGCCTGTTTT TGAATGTTAT	9540
CATACTCAAT TCCCAGTTCT TGCTCGACAG TAGTGATATG TTTTGAATG GTTTCTAAAT	9600
CATGACTCTT CTGTTTTCTT TTTTCAAGGA TACGAACCAA GTGACTGCGG ATGCCTTCCT	9660
CAGCGAAAAA GAGGCTGTG TCAAAGATCT TGGTATCAAT CTGCTGAACC TTGTCTTCTT	9720
CGATCAGGTA GGAGAGCTCT TGGGCAACTT GGCTGGGGTC TAGTTCCACG GGACGGGAAG	9780
ACTCAAGGAG AGTAAGGGTT TGTTCAGCA AATCCCGTGC TTCAACATAG GTGTCCCCTG	9840
TTTCCATACA GGCCTGAAAA AGACTGTGAA CTAGACCGGC GCGGAAGCGT TCAGGAGCCT	9900
GACTTTTCGAT GCCTAGTTCC TCAGCTAGTT GGTACGCAAT GGTAAAGCCC AAACCCTTGA	9960
TATCCTCAAC CAGCTGGTAG GGATAATTTT CAACCACATC AAGGGTTTCT TCCTTGTAAT	10020
AGTCTTGAAT CTGAAAGGCT AGTTTGTGG GAATGCCGTA GTTGGCTAGT TTGGCCAAAA	10080
TCATCTCCGT TCCGTAGTTG AGACGGAGAG TGGAGACGAA AGCCTCGCGA TTTTGGCAG	10140
AGAGTCCTGC CATGCCTTCT AACTTTTCTG GGTGTTGCAA AATTTCGTCA ATGGTATTTT	10200
CGCCATAGGT ATCCACGATT TTCTGAGCTG TCTTGAGACC AATCCCCTTG AAATGGCTAC	10260
TTGAAAAGTA CTTGACCAAG CCCTTACTAG TTGGTTTTCG GCGATCATAA CGACTGATTT	10320
GCAGTTGTTT TCCATACCTG GAGTGCTGGA CAATTTGCC CCAAAAAGTA TAGTCTTCGC	10380



657

CCTCAATTAC ATCAGCCATG GTTCCTGTGA CAATGATTTC AAAATCATCA AAATCCTCTG	10440
CGTCCGTATC GTCGATTCTT AGGAGGAGGA TGCATAAAAA ATTGCTGGGA TTTTCAAAAA	10500
TAATCCGTTC AATAGTTCCT GAAAAATAAA CTTCCATAAA ATTCCTTTGC ATGAATAGGT	10560
GAGAGTTGGG ATTGTTTTTA TTTTATACTC TTCGAAAATA TCTTCAAACC ACGTCAGCTT	10620
CCATCTGCAA CCTCAAAACA GTATTTTGAG CTGACTTCGT CAGTTCTATC CACAACCTCA	10680
AAACACTGTT TTAAGCAGCC TACGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTCATTGAG	10740
TATTTGTAAA TAAACAATCA CTTCTCACGA TAGAAGAAGA GGCTGAGATT GGTGATTCTC	10800
TGCCTCTTAG GTTCTTAAA ATGTTCCGAT ACGGGTGATT GGCCATAAGC GGAATTTAGC	10860
TTCCCTGTG ATATCTTTTG CTTTGAAGGT ACCTACGTGG CGGCTGTCGC TCGAAACCAA	10920
GCGGTCACTC CCGAGGAGAA GGTATTCTCC TTCTGGAACA GTAAAGCTAA AGTTGGTGTT	10980
GTAGTTGACA TCAACTGTGA AGGCTTGAGC TTTTGTAGCG ATACTTCTAA AGAAAGTTCC	11040
TTTATTTCTT TCAAAGCCCT TGCCTGAGTA AGTGCTTTGG AGTTTGTCTAT CCTTGAAGCG	11100
TTTGATATAG TCTGCTAGAT AAGGCTCGTC CGTTTCTTTG TCATTGATGT AGAGTTTATC	11160
ATTTTCGTAA CGAATGGTGT CGCCAGGCAT TCCAATCACG CGCTTGACGA TGTCCTTATT	11220
GCCATCTTCC TCATGGGCCA CCACGATATC AAAACGGTCA ATAGGAAGGT CTTTTACAAC	11280
GAAGAGAATT TCGCCATCCG CTAGGGTCGG ATCCATGGAA TGTCTTCTA CGCGAACATT	11340
GCTCCAAAAA AAGATACGAC TTAAAGCTAG TAATGACAGA ATTAGGAGGA ACAATCCCCA	11400
CTCTTTTAAG AAATTTTAA ATGAATTCAT AACTTACCTT TCTAAGCGTT TTTTCGCTTT	11460
TTCAGTGTTT TTAAAGTGCA ATTTGGCGCA GAAGCTGAGT CCCTGCATAC CATAGGCTTG	11520
CAAAATCTGG CTAGCCACCT TGTCAGAAGC CGTTCCAGCT CCACTTGGGA GCTGATAACC	11580
CAGTTCTCGT CCCAAATTTT CAAGATTTTC CAGAAAGAGA TCACGCGCAA TGACAGAAGA	11640
AACTGCGACA GACAAGTATT TGCCCTCAGC CTTTCTTCTT AAGCTGATAG GATTGCTGAA	11700
ACGATTGGCC TCTTGTGCCA AGTACTTGTC ATAATTTTTA GCACTGGTAA AGGCATCAAT	11760
CACAATTTTC TCAGGCTGAA CACCTTTTTG AAGGAGGAGA TAGATAGCCT GATTATGGAG	11820
GGCAACCTTA ACCGAAACAG CGTTGTAGCG GTCTCCGATG ACCTCGTTGT ACTTGCTGGG	11880
TGAGAGAAGG AGTGCTGGT GCTGAATTTT TTCCTTGAGA ATAGGAGTAA TCTGACGGAT	11940
CTTTTGGTCG GTCAGAGTCT TAGAATCCCC CACACCGAGT TTTTCGTAATA AGTCGTGCTG	12000
GTCAGGTGTG ACAAAGGCAG CCACAACCTGC AAGCCCCACCA AAGTAGGAAC CATTTCCCAC	12060
CTCATCTGTC CCAATTAAAG GAAGATTTTG TCCGCTGGTT TGCTCTACAG CTTGATAGCC	12120

658

AAAGAACTG GCGTATTTT CAGCCCCTC ACCCTGAAGC AAGATTTTC CAGAAGTATA	12180
GATAGAAACC GTTGCTGAG GTAGTTTCAA AAAGTAGCGG ATATAGGGAT TCTTGCTAGG	12240
AGCCAGACTG GTTTGATAGT GTTCAAGAAA AGCCTGAATA TCCTTTTCGC TTGGTGTGAG	12300
TGTGATACTT GCCATAGTTT CTATTGTACC ACAAAGCAG TAAAATTTGT AAAAAGTAC	12360
AAAAATAGCG AATTTTGGTA TAATATCGTG AGGTGAATTT TATGGCAAAT CTAAATCGAT	12420
TCAAATTTAC ATTCGGGAAA AAATCGTTAA CCTTGACAAG CGAACATGAC AACCTTTTA	12480
TGGAGGAAAT CGCTAAGGTT GCGACAGAAA AATACCAAGC AATTAAAGAA CAAATGCCTA	12540
GCGCAGATGA TGAACAATC GCTCTTTGT TGGCAGTCAA CTGTTTATCA ACTCAGCTCA	12600
GCCGTGAGAT TGAATTTGAC GATAAGGAGC AAGAGCTAGA AGAACTCCGT CACAAGCTTG	12660
TGACTTGTA GCAAGAACAG AGCAAGATTG AGGATTCCTT ATGATTTTAT TCCTTCTTCT	12720
ATTGGTCTTG GTTTGGGGAT TTTATATCGG CTATCGGAGA GGCCTGCTCT TACAGGTTTA	12780
TTACCTGATT TCAGCCATGG CATCGGCTTT TATGGCTGGC CAGTTTATA AGGGGCTTGG	12840
AGAGCAATTC CATTTATTGC TCCCTTATGC AAATTCGCAG GAAGGTCAGG GGACTTTCTT	12900
TTTCCCATCG GATCAACTCT TTCAGCTGGA TAAGGTCTTT TATGCAGGTA TCGGCTACTT	12960
GCTTGATTTT GGGATTGTCT ATAGCATTTG TCGTTTACTT GGTCTTCTCT TACACTTGAT	13020
TCCTAGCAAA AACTGGGTG GTAAGTTGTT CCAAGTTTCA GCAGGTATCT TGTCCATGTT	13080
GGTGACCTTA TTTGTCTTGC AAATGGCCTT GACAATCTTG GCGACCATCC CCATGGCAGT	13140
TATACAAAAT CCTCTTGAAA AGAGTATCGT CGCAAAACAC ATCATCCAGA GCATACCGGT	13200
AACAACCACT TGGCTCAAAC AAATCTGGGT GACAAATTTA ATCGGATAAA AAGGGCAGGA	13260
GTTCCTTAG CCCTTTGTTT ACAGATTGTA CTCGAATCTA TCAGAATGTA AAAAGTACC	13320
ACACCTAGAC ATTCAAAGAC AAGGAAATAA AGATGAATAA GAAAATATTA GAAACATTAG	13380
AGTTCGATAA GGTCAAGGCC TTGTTTGAGC CTCATTGTGTT GACCGAGCAG GGCTTGAGC	13440
AATTGAGACA ACTGGCTCCG ACTGCCAAAG CAGATAAAAT CAAACAGGCT TTTGCTGAGA	13500
TGAAGGAAAT GCAGGCTCTT TTCGTCGAGC AACCGCATTT TACTATTCTC TCAACTAAGG	13560
AAATTGCAGG AGTCTGCAAG AGGTTGGAGA TGGGAGCGGA TCTCAATATC GAGGAGTTCC	13620
TACTCTTGAA ACGCGTGCTT CTTGCCAGCC GAGAACTTCA AAATTTTAC ACCAATCTGG	13680
AAAATGTCAG CTTGGAAGAA TTAGCCCTTT GGTGAGAGAA ATTACATGAT TTTCCGCAAT	13740
TACAAGGAAA TCTTCAGGCC TTTAATGATG CGGGTTTCAT TGAATTTT GCCAGTGAAG	13800
AATTGGCGCG AATCCGTCGA AAAATACATG ATAGCGAGAG TCAGGTACGC GATGTTTAC	13860
AAGACTTGCT CAAGCAAAAA GCGCAGCTGT TGACGGAAGG AATTGTTGCT AGCAGAAATG	13920

559

CCCCTCAGGT TTTACCAAGT AAAAACACCT ACCGCAATAA GATTGCAGGT GTCGTTTCATG	13980
ATATTTCTGC TAGTGGAAC ACCGTCTATA TCGAACCCCG TGAGGTAGTC AAATGAGCG	14040
AAGAAATTGC TAGTCTGCGA GCAGATGAGC GCTATGAAAT GCTTCGCATT CTCCAAGAAA	14100
TTTCTGAGCG TGTCCGCCCT CATGCGGCTG AGATTGCTAA TGACGCTTGG ATTATCGGTC	14160
ATCTGGACTT GATTTCGTGCC AAGGTTTCAT TTATCCAAGA AAGACAAGCA GTCGTGCCTC	14220
AGCTGTGAGA AAATCAAGAG ATTCAACTGC TCCATGTCTG CCATCCTTTG GTCAAAAATG	14280
CCGTCGCAAA TGATGTCTAT TTTGGTCAAG ATTTAACAGC TATTGTCATT ACAGGTCCCA	14340
ATACAGGTGG GAAGACCATC ATGCTCAAAA CTCTGGGCTT GACACAGGTC ATGGCCAGT	14400
CAGGATTGCC GATTTTAGCA GACAAGGAA GTCGTGTTGG TATTTTGA GAAATCTTTG	14460
CTGATATTGG AGATGAGCAG TCTATTGAGC AGAGCTTGTG TACCTTCTCT AGTCATATGA	14520
CCAATATCGT GGATATTCTT GGCAAGGTCA ACCAACATTC ACTCTTACTT TTGGATGAGT	14580
TGGGGGCTGG TACTGATCCC CAAGAGGGAG CAGCCCTTGC CATGGCTATT CTGGAGGACC	14640
TTCCCTGCG TCAAATCAAG ACCATGGCGA CGACCCACTA TCCAGAACTC AAGGCCTACG	14700
GTATTGAGAC AGCCTTTGTG CAAAATGCCA GTATGGAGTT TGATACTGCA ACTCTTCGCC	14760
CGACCTATCG CTTTATGCAG GGTGTTCTG GCCGAAGTAA TGCCTTTGAA ATTGCCAAAC	14820
GTCTAGGCCT ATCTGAAGTT ATCGTAGGAG ATGCCAGTCA GCAGATCGAT CAGGACAATG	14880
ACGTCAATCG TATCATTGAG CAATTAGAAG AGCAGACGCT GGAAAGCCGC AAACGTTTGG	14940
ACAATATCCG TGAGGTGGAG CAAGAAAATC TCAAGATGAA CCGTGGCGTA AAAAACTCT	15000
ACAACGAGCT TAATCGTGAA AAGGAAACCG AGCTTAACAA GGCGCGTGAA CAGGCTGCTG	15060
AGATTGTGGA TATGGCCCTA AGTGAAAGTG ACCAGATTCT CAAAAATCTC CACAGTAAAT	15120
CCCAACTCAA GCCCCACGAA ATCATTGAAG CCAAGGCCAA GTTGAAAAAA TTGGCTCCTG	15180
AAAAAGTGGA CTTGTCTAAA AATAAGGTCC TTCAAAGGC CAAGAAAAAA CGAGCTCCAA	15240
AGGTGGGAGA TGATATCGTG GTTCTCAGTT ATGGTCAGCG TGGTACCTTG ACCAGTCAAC	15300
TCAAGGACGG TCGCTGGGAA GCCCAAGTTG GCTTGATTAA GATGACCTTG GAAGAGAAAG	15360
AGTTTGATCT TGTCAAGCC CAGCAAGAAA AACCAGTCAA GAAGAAACAG GTCAATGTTG	15420
TGAAACGAAC TTCTGGGCGA GGACCTCAAG CTAGACTGGA TCTTCGAGGC AAGCGCTATG	15480
AAGAAGCCAT GAATGAGCTA GATACCTTCA TCGACCAAGC CTTGCTTAAC AATATGGCTC	15540
AAGTTGATAT CATCCATGGT ATCGGAACAG GAGTCATCCG TGAAGGAGTT ACCAAATACT	15600
TGCAAAGAAA CAAACATGTC AAGAGTTTCG GCTATGCCCC ACAAATGCT GGAGGCAGTG	15660

660

GTGCGACTAT TGTCACCTTTT AAAGGATAGC AGTATTCTGG ACTTTATAAA GTAAAACTG	15720
TTGAACTAAT TTTTACTAAT AAACACATTG ACAAAGCCA ACATTTTTTG TAAAATTAGA	15780
ATCAATTAAA TACCAACACC GAATGAAGTT TAATAGAAGT GGGGAATCGT TTGATTTTCC	15840
ATGACTGTAA ATGGACGGAA CTCTGGAGAG ACCGTAAAGG CACCGAAGGG CAAGGCAGGC	15900
AACTGCTCAA ACTCTCAGGT AAAAGGACAG AGCTAGGATA GATCGCTTTT TAGCATTTAT	15960
CTAAGCATTC CAGAGTACAT GTATCTTGCA TGTGCTCTTT CTTTTGGGGT TGAAACGATA	16020
GGAGAAGGAA ATGTTAGAAT TGCTTAAATC AATCGATGCT TTTGCTTGGG GACCGCCCT	16080
CTTGATTTTA TTGGTCGGAA CAGGGATTGA CCTAACTATT CGGCTAGGAC TCTTGACAGT	16140
TTTGGCTCTA CCCAAGGCCT TTCAGCTTAT TTTTATCCAG GATAAGGGAC ATGGTGATGT	16200
ATCCAGTTT GCAGCTCTGT GTACAGCCTT GGCATCAACT GTTGAACAG GAAATATCAT	16260
AGGAGTTGCG ACGGCTATCA AGGTTGGTGG ACCAGGAGCT CTATTTTGGA TGTGGATGGC	16320
GGCTTTCTTT GGAATGGCTA CCAAGTATGC GGAAGGACTC TTGGCCATCA AATACCGCAC	16380
CAAGGACGAC CATGGTGCAG TAGCGGGAGG TCCCATGCAT TATATCCTTC TAGGGATGGG	16440
AGAAAAGTGG CGACCACTTG CTGTTTTGTT TGCAGTAGCA GGAGTATTGG TTGCTCTCTT	16500
GGGAATCGGA ACCTTCACCC AAGTCAACTC GATTGCAGAA TCTATCCAAA ATACAACGAC	16560
GATTTCGCCA GCCATCACAG CTCTCGTCTT GTCTGTCTTT GTAGCGATTG CAGTCTTTGG	16620
TGGACTCAAG TCTATTTCTA AGGTTTCAAC TACTGTTGTT CCTTTTATGG CCATCATTTA	16680
TATCTTAGGA ACTCTTACAG TTATTTTCTT TAATATCGGA AAAATCCCTG GCACAATCGC	16740
TTTAGTCTTT ACCTCAGCTT TTAGTCCCCT TGCTGCGGTA GGTGGATTG CTGGTGCTAG	16800
CGTTCCGATG GCTATTCAAA ATGGTGTGGC GCGTGGTGTG TTCTCAAACG AATCTGCTCT	16860
GGGTTCTGCT CCTATTGCAG CTGCAGCTGC CAAGACAAAT GAACCAGTAG AGCAAGGTTT	16920
GATTTCATG ACAGGAACCT TTATTGATAC CCTCATCATT TGTACTCTAA CTGGTTTGAC	16980
CATCTTGGTA ACTGG	16995

## (2) INFORMATION FOR SEQ ID NO: 83:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28473 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

CCGGGGCTTT TGTAGTATAA TAGAGATACG TTTTGAAAGT AGGAGGTATC TATGGACTTA

60

661

ACTAAGCGCT TTAATAAACA GTTAGATAAA ATTCAAGTTT CGTTCATTCC TCAGTTTGAC	120
CAGGCTATTT CGGAGATTCC TGGGGTCTTG CGTTTGACCT TGGGGGAACC TGATTTTACA	180
ACGCCAGACC ATGTCAAGGA GCGGGCAAG CGAGCGATTG ATCAGAACCA ATCCTACTAT	240
ACAGGGATGA GTGGTCTGCT GACTCTACGT CAGGCAGCCA GTGACTTTGT TAAGGAAAAG	300
TACCAACTGG ACTATGCTCC TGAAAATGAA ATCTTGTTA CAATTGGGGC GACAGAGGCT	360
TTATCTGCGA CTTTGACGGC TATTTTGAA GAGGGAGACA AGGTACTTTT GCCAGCTCCT	420
GCTTATCCAG GCTATGAACC GATTGTAAAC TTAGTTGGGG CAGAAATTGT TGAGATTGAT	480
ACGACTGAAA ATGGTTTTGT CTTGACTCCT GAGATGTTGG AGAAGGCCAT TTTGGAGCAG	540
GGTGATAAGC TCAAGGCGGT TATTCTCAAC TATCCAGCCA ATCCGACAGG AATTACCTAC	600
AGTCGAGAGC AGTTAGAGGC CTTGGCAGCT GTTTTACGCA AGTACGAAAT TTTTGTGTGTC	660
TGTGATGAGG TTTACTCAGA ATTGACCTAC ACAGGCGAAG CCATGTGTCT CTAGGAACGA	720
TGTTGAGAGA CCAGGCTATT ATTATCAATG GTTTGTCTAA ATCGCATGCC ATGACAGGTT	780
GGCGTTTGGG GCTGATTTTC GCTCCTGCGA CCTTCACAGC CCAGTTAATC AAGAGTCACC	840
AGTACTTGGT CACTGCCGCA AATACCATGG CGCAACATGC TGCGGTAGAA GCCTTGACGG	900
CTGGTAAAAA CGATGCGGAC CCATGAAGAA GGAATATATC CAACGTCGGG ACTATATCAT	960
CGAAAAATG ACTGCTCTTG GTTTTGAGAT TATCAAACCA GACGGTGCCT TCTATATTTT	1020
TGCTAAATTT CCAGCGGGCT ACAATCAAGA CTCCTTTGCT TTTCTGAAGG ATTTTGCTCA	1080
GAAGAAGGCC GTTGCCTTTA TCCCTGGTGC AGCCTTTGGA CGTTACGGGG AAGGCTACGT	1140
CCGCCTATCT TATGCAGCCA GCATGGAGAC TATCAAAGAA GCCATGAAAC GACTTGAGGA	1200
GTACATGAGA GAAGCATGAT TCACTCTATC ACGAGTCAAG GCTTGGTGCT TTACAATCGC	1260
AATTTTCGTG AGGATGACAA GCTCGTCAAA ATTTTACAG AGCAGGTTGG CAAACGCATG	1320
TTTTTTGTCA AACACGCTGG TCACTCTAAG CTGGCGCCTG TTATTCAGCC CTTGGTGCTG	1380
GCACGATTTT TCTTGCGAAT CAATGATGAC GGACTCAGTT ACATCGAAGA CTATCATGAG	1440
GTCATGACTT TTCCCAAGAT TAATAGTGAC CTCTTTGTCA TGGCCTATGC GACCTATGTG	1500
GCAGCTCTTG CAGATGCTAG TTTGCAGGAC AATCAGCAGG ATGCTCCCTT GTTTGCTTTT	1560
TTGCAAAAGA CTTTGAGTT GATGGAAGCA GGCTTGGATT ATCAGGTTTT GACCAATATT	1620
TTTGAAATTC AAATTTTGAC TCGATTTGGA ATCAGCCTCA ATTTTAAATGA GTGTGTCTTC	1680
TGCCATCGGG TTGGTCAGGC TTTTGACTTT TCTTTCAAAT ATGGAGCCTG CCTCTGTCCA	1740
GAGCATTATC ATGAGGATAA GAGACGTTGT CATCTCAATC CCAATATCCC CTATCTGCTC	1800

662

AATCAATTTTC	AAGCTATTGA	TTTTGAGACT	TTGGAGACCA	TTTCGCTCAA	GCCTGGAATC	1860
AAGCAAGAGC	TACGCCAATT	TATGGATCAA	TTATATGAAG	AGTACGTTGG	GATTCACCTA	1920
AAATCAAAGA	AATTTATTGA	TTCCCTAGCA	GACTGGGGAC	AATTACTAAA	AGAGGAAAAG	1980
AAATGAAAAA	AATCGCAGTA	GATGCCATGG	GGGGCGATTA	CGCACCTCAG	GCCATTGTTG	2040
AGGGTGTCAA	TCAAGCCCTA	TCTGACTTTT	CAGATATCGA	GGTTCAACTT	TACGGAGATG	2100
AAGCTAAAAT	CAAGCAATAT	CTGACAGCGA	CAGAGCGCGT	CAGCATTATC	CATACGGATG	2160
AGAAGATTGA	TTCGGATGAT	GAACCTACGA	GAGCTATTTC	GAATAAGAAA	AATGCCAGTA	2220
TGGTATTGGC	AGCCAAGGCT	GTCAAAGATG	GTGAAGCAGA	CGCTGTCCTT	TCGGCTGGGA	2280
ATACAGGTGC	CTTGTGGCA	GCAGGATTCT	TCATCGTGGG	TCGTATCAAG	AATATCGACC	2340
GTCTGGACT	CATGTCTACC	TTGCCTACCG	TTGATGGAAA	AGGTTTTGAC	ATGCTAGACC	2400
TTGGTGCCAA	TGCAGAAAAT	ACAGCCAGC	ACCTCCATCA	ATATGCGGTT	CTAGGTTCTT	2460
TCTATGCTAA	AAATGTCCGT	GGCATTGCGC	AACCACGCGT	TGGTTTGCTC	AACAACGGAA	2520
CAGAGAGTAG	CAAGGGCGAC	CCGCTTCGTA	AGGAAACTTA	TGAATTACTG	GCGGCTGATG	2580
AAAGTTTGAA	CTTTATCGGA	AACGTGGAAG	CGCGTGATTT	GATGAATGGC	GTTGCAGATG	2640
TTGTTGTGGC	AGATGGTTTC	ACGGGAAACG	CTGTGCTCAA	ATCCATCGAA	GGGACAGCTA	2700
TGGGAATCAT	GGGCTTGCTC	AAGACAGCTA	TTACAGGTGG	TGGTCTTCGA	GCGAAACTAG	2760
GTGCCCTCCT	TCTCAAGGAC	AGCCTCAGTG	GTTTGAAAAA	ACAGCTCAAT	TATTCAGATG	2820
TTGGTGGAGC	GGTCTTGTTT	GGTGTTAAGG	CACCTGTTGT	CAAGACTCAT	GGCTCAAGCG	2880
ATGCCAAGGC	TGTTTATAGT	ACGATTCTGC	AGATCCGTAC	CATGCTAGAA	ACAGACGTGG	2940
TTGCCCAGAC	TGCCCGTGAA	TTTTCAGGAG	AATAAAAGAG	ATGACAGAAA	AAGAAATTTT	3000
TGACCGTATT	GTGACCATTA	TCCAAGAGCG	ACAGGGAGAG	GACTTTGTCT	TGACAGAATC	3060
CTTGAGTCTG	AAAGACGATT	TGGATGCGGA	TTCTGTTGAC	TTGATGGAGT	TTATCTTGAC	3120
TCTGGAAGAT	GAATTTAGTA	TCGAAATCAG	CGATGAAGAA	ATTGACCAAC	TCCAAAACG	3180
AGGAGATGTG	GTTAAAATCA	TTCAAGGAAA	ATAGCAATCG	GAGTTCCAAG	TCAACGGAAG	3240
TAGATGGTTT	TTAGAAATGA	GAAATATCGG	ACAAGCTGGT	AAAATCTTGG	CTGACAGTGG	3300
TTATCAAGGG	CTCATGAAGA	TATATCCTCA	AGCACAAACT	CCACGTAAAT	CCAGCAAACCT	3360
CAAGCCGCTA	ACAGTTGAAG	ATAAAGCCTG	TAATCATGCG	CTATCTAAGG	AGATAAGCAA	3420
GGTTGAGAAT	ATCTTTGCCA	AAGTAAAAAC	GTTTAAAAATG	TTTTCAACAA	CCTATCGAAA	3480
TCATCGTAAA	CGCTTCGGAT	TACGAATGAA	TTTGATTGCT	GGTATTATCA	ATCATGAACT	3540
AGGATTCTAG	TTTGCAGGA	AGTCTAATAG	TAAAAAAGTG	ATTAGAAAAC	ATCTTTTTTA	3600

663

AAAATAGAGA TGATTTTGAA ACAAAAAAGC TAATTCAAGA CGTTTCGATG CCAATTCAAG	3660
ATTTGGATGA AAAAAATTAA TAGATACTGT TATACTAAAC TTGTCAAGTT TGTAACAAGA	3720
CAAAATATTAA AAATAAAAAA GAGGTATTCC TTATGAATAC AAAAACGATG TCACAATTTG	3780
AAATTATGGA TACTGAGATG CTTGCTTGCG TTGAAGGTGG CGGATGCAAT TGGGGAGATT	3840
TTGCCAAAGC AGGTGTTGGA GGAGGAGCAG CACGAGGTCT TCAGCTAGGA ATTAAAACAA	3900
GAACATGGCA AGGTGCAGCA ACTGGTGCTG TGGGAGGAGC TATACTTGGA GGTGTGGCCT	3960
ATGCAGCGAC ATGTTGGTGG TAATTATGGA TTTTAAAAGT TTTATTATTG GTTTAGTAGT	4020
TGGTATATTT GGTCTTATA TGGATGATT AATTAGAAAA AAATTTTAA AGTCTTCGGA	4080
GAAGAAAACA GAAAAATCTG TTAATAATA ATCAAACTA TAAATGATGA ATCTGAATCA	4140
AAATTATTTT GCGCATGTAA AGAGGAGTCT TATAGTAACG AGTCAAAAAA GGAGTAACTA	4200
TGAATCGTAA TTTAGAACGG TGTATCTAT TCTGACTAGG AATAGATCAT ACCAGAGGTA	4260
GCTTAGAAAT AGCAGAGACA TTAGAAATTG AAGTAATAAA TAGGATGTCG TAAGTGTTAC	4320
TATCAATGAT TTATTTGTTT CAAGCTTGCC TAGGGTGACA GTAAAAATC AATTCCTTT	4380
CAATAGCATA TTTTCTAGTG GCAGGACTCT TGTCTGCCT ATTTTTTTAT CCAAAAAGTG	4440
CAGTTGGGAG GGAGATAGGC TCATTGCGA AGGAAGTCCA GTTTTTGTTT AGTGATTGGG	4500
GTAAGATAGT TGTTATCAGA TGAGTTAATA CTCTTCGAAA ATCAAATTCA AACCACGTCA	4560
ACGTCGCCTT GCCGTATATA TGTGACTGAC TTCGTCAGTC CTATCTACAA CCTCAAAACA	4620
GTGTTTTGAG CAGCCTACGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA	4680
GGGAAAAGGA CATGAATATG AAATTGCGA AACGTCATTA TCGTCCGAG GTGGATCAGA	4740
TGGACTGCGG TGTAAGTTCA TTAGCCATGG TTTTGGCTA CTATGGTAGT TATTATTTTT	4800
TGGCTCACTT GCGAGAATTG GCTAAGACGA CCATGGATGG GACGACGGCT TTGGGCTTGG	4860
TCAAGGTGGC AGAGGAGATT GGTTTTGAGA CGCGAGCCAT TAAGGCAGAT ATGACGCTTT	4920
TTGACTTGCC GGATTTAACT TTTCTTTTG TTGCCCATGT GCTTAAGGAA GGGAAATTGC	4980
TCCACTACTA TGTGGTGA CTGGCAGGATA AGGATAGCAT TCATATTGCC GATCCAGATC	5040
CCGGGGTGAA GTTGACTAAA CTGCCACGTG AGCGTTTTGA GGAAGAATGG ACAGGAGTGA	5100
CTCTTTTTAT GGCACCTAGT CCAGACTATA AGCCTCATAA GGAACAAAAA AATGGTCTGC	5160
TCTCTTTTAT CCTATATTA GTGAAGCAGC GTGGCTTGAT TGCCAATATC GTTTTGGCAA	5220
CACTCTTGGT AACCGTGATT AACATTGTGG GTTCTTATTA TCTGCAGTCT ATCATTGATA	5280
CCTATGTGCC AGATCAGATG CGTTCGACAC TAGGGATTAT TTCTATTGGG CTAGTCATCG	5340

664

TCTACATCTT CCAGCAAATC TTGTCTTACG CTCAGGAGTA TCTCTTGCTT GTTTGGGGC	5400
AACGCTTGTC GATTGACGTG ATTTTGTCCT ATATCAAGCA TGTTTTCAC CTCCTATGT	5460
CCTTCTTGC GACACGCAGG ACAGGGGAGA TCGTGTCTCG TTTTACAGAT GCTAACAGTA	5520
TCATCGATGC GCTGGCTTCG ACCATCCTTT CGATTTTCCT AGATGTGTCA ACGGTGTCA	5580
TTATTTCCCT TGTTCATTT TCACAAAATA CCAATCTCTT TTTTCATGACT TTATTGGCGC	5640
TTCTATCTA CACAGTGATT ATCTTTGCCT TTATGAAGCC GTTTGAAAAG ATGAATCGGG	5700
ATACCATGGA AGCCAATGCG GTTCTGTCTT CTTCTATCAT TGAGGACATC AACGGTATTG	5760
AGACTATCAA GTCCTTGACC AGTGAAAGTC AGCGTTACCA AAAAATTGAC AAGGAATTTG	5820
TGGATTATCT GAAGAAATCC TTTACCTATA GTCGAGCAGA GAGTCAGCAA AAGGCTCTGA	5880
AAAAGGTTGC CCATCTCTTG CTTAATGTCT GCATTCTCTG GATGGGGGCT GTTCTGGTCA	5940
TGGATGGCAA GATGAGTTG GGGCAGTTGA TTACCTATAA TACCTTGCTG GTTTACTTTA	6000
CTAATCCTTT GGAAAATATC ATCAATCTGC AAACCAAGCT TCAGACAGCG CAGGTGCCA	6060
ATAACCGTCT AAATGAAGTG TATCTAGTAG CTTCTGAGTT TGAGGAGAAG AAAACAGTTG	6120
AGGATTTGAG CTTGATGAAG GGAGATATGA CCTTCAAGCA GGTTCAATTAC AAGTATGGCT	6180
ATGGTCGACA TGTCTTATCG GATATCAATT TAACCGTTCC CCAAGGCTCT AAGGTGGCTT	6240
TTGTGGGGAT TTCAGGGTCA GGTAAGACGA CTTTGCCAA GATGATGGTT AATTTTACG	6300
ACCCAAGTCA AGGGGAGATT AGTCTGGGTA GTGTCAATCT CAATCAGATT GATAAAAAAG	6360
CCCTGCGCCA GTACATCAAC TATCTGTCTC AACAGCCCTA TGTCTTTAAC GGAACGATTT	6420
TGGAGAATCT TCTTTTGGGA GCCAAGGAGG GGACGACACA GGAAGATATC TTACGGCGCG	6480
TCGAATTGGC AGAGATTGCA GAGGATATCG AGCGCATGCC ACTGAATTAC CAGACAGAAT	6540
TGACTTCGGA TGGGGCAGGG ATTTTCAGGTG GTCAACGTCA GAGAATCGCT TTGGCGCGTG	6600
CTCTCTTGAC AGATGCGCCG GTCTTGATTT TGGATGAGGC GACTAGCAGT TTGGATATTT	6660
TGACAGAGAA GCGGATTGTC GATAATCTCA TTGCTTTGGA CAAGACCTTG ATTTTCATTG	6720
CTCACCGCTT GACTATTGCT GAGCGGACAG AGAAGGTAGT TGTCTTGAT CAGGGCAAGA	6780
TTGTGAAGA AGGAAAGCAT GCTGATTTGC TTGCACAGGG TGGCTTTTAC GCCCATTTGG	6840
TCAATAGCTA GAAAGAGGAG AGGATGAAAC CAGAATTTT AGAAAGTGCG GAGTTTTATA	6900
ATCGTCGTTA CCATAATTTT TCCAGTAGTG TGATTGTACC CATGGCCCTT CTGCTTGTGT	6960
TTTTACTTGG CTTTGCAACT GTTGACAGAGA AGGAGATGAG TTTGTCCACT AGAGCTACTG	7020
TCGAACCTAG TCGTATCCTT GCAAATATCC AGTCAACTAG CAACAATCGT ATTCTTGTC	7080
ATCATTTGGA AGAAAATAAG CTGGTTAAGA AGGGGATCT TTTGGTTCAA TACCAAGAAG	7140



665

GGGCAGAGGG TGTCCAAGCG GAGTCCTATG CCAGTCAGTT GGACATGCTA AAGGATCAAA	7200
AAAAGCAATT GGAGTATCTG CAAAAGAGCC TGCAAGAAGG GGAGAACCAC TTTCCAGAGG	7260
AGGATAAGTT TGGCTACCAA GCCACCTTTC GCGACTACAT CAGTCAAGCA GGCAGTCTTA	7320
GGGCTAGTAC ATCGCAACAA AATGAGACCA TCGCGTCCCA GAATGCAGCA GCTAGCCAAA	7380
CCCAAGCCGA AATCGGCAAC CTCATCAGTC AAACAGAGGC TAAAATTTCG CATTACCAGA	7440
CAGCTAAGTC AGCTATTGAA ACAGGTGCTT CTTGGGCCGG TCAGAATCTA GCCTACTCTC	7500
TTTACCAGTC CTACAAGTCT CAGGGCGAGG AAAATCCCCA AACTAAGGTT CAGGCAGTTG	7560
CACAGGTTGA AGCACAGATT TCTCAGTTAG AATCTAGTCT TGCTACTTAC CGTGTCCAGT	7620
ATGCAGGTTT AGGTACCCAG CAAGCCTATG CGTCAGGTTT AAGCAGTCAA TTGGAATCCC	7680
TTAAATCCCA ACACTTGGCA AAGGTTGGTC AGGAATTGAC CCTTCTAGCC CAGAAAATTT	7740
TGGAGGCAGA GTCAGGTAAG AAGGTACAGG GAAATCTTTT AGACAAGGGG AAAGTTACGG	7800
CGAGTGAGGA TGGGGTGCTT CATCTTAATC CTGAGACCAG TGATTCTAGC ATGGTTGCAG	7860
AAGGTGCCCT ACTAGCCCAA CTTTATCCAT CTTTGGAAAG AGAAGGGAAA GCCAACTCA	7920
CAGCTTATCT AAGTTCAAAA TATGTAGCAA GAATCAAGGT CGGTGATTCT GTTCGCTATA	7980
CTACGACTCA TGATGCCGGG AATCAACTTT TCCTAGATTC TACTATTACA AGTATTGATG	8040
CGACAGCTAC TAAGACTGAG AAAGGGAATT TCTTTAAAT CGAGGCCGAG ACTAATCTAA	8100
CTTCGGAGCA GGCTGAAAA CTTAGGTACG GGGTGGAAAG CCGCTGCGAG ATGATTACGG	8160
GCAAGAAAAG TTACCTACGT TATTATTTGG ATCAATTTTT GAACAAAGAG TAATGTTCTG	8220
GTTTTTAGAG TTAAATAATT TTTAACTGT GAGAAAGATT CTTCTTGCAG TTTTTCTTT	8280
ACAATTTTTG AAAACATCT ACTATTTATT CGGTAAAT CTTGTGTTTT TTGGTTTTTT	8340
GTGGTAAAT GTGCTCAAGT AATACGAAAG GCGAACTTTA AAATGTCAA ACAATTGATC	8400
TATTCGGGAA AAGCTAAAGA TATCTATACA ACTGAGGATG AAAATCTTAT TATTTCAACT	8460
TACAAGGACC AGGCGACTGC TTTCAACGGT GTCAAGAAGG AGCAGATTGC AGGTAAGGGA	8520
GTCTTGAATA ATCAGATCTC ATCTTTTATT TTTGAGAAAT TAAATGTGGC TGGTGTGGCG	8580
ACTCACTTTG TGGAGAACT TTCAGACACG GAACAACCTA ATAAAAAGGT TAAGATTATT	8640
CCTTTGGAAG TCGTGCTCCG CAACTATACT GCTGGTTCCT TTTCAAAACG TTTTGGTGTG	8700
GATGAGGGAA TCGCCTTGGG GACTCCGATT GTCGAATTTT ACTACAAAA TGATGATTTG	8760
GATGATCCAT TTATCAATGA TGAGCATGTG AAATTCCTAC AGATTGCGGG TGACCAGCAG	8820
ATTGCCTACT TGAAGGAAGA AACGCGTCGT ATCAATGAAC TATTGAAAGT CTGGTTTGCT	8880

666

GAGATTGGGC TTAAATTGAT TGACTTTAAG CTAGAGTTCC GTTTGTGACAA GGATGGCAAG	8940
ATTATCTTGG CAGACGAATT TTCACCAGAT AACTGCCGCT TGTGGGACGC TGATGGCAAC	9000
CACATGGATA AGGATGTTTT CCGTAGAGGA TTGGGAGAAC TAACCGACGT TTATGAGATT	9060
GTTTGGGAAA AGTTGCAGGA ATTGAAATAA TCTGTTTGCA ACGGAAAACC TTCGTCTCTC	9120
AACTAAAAGG ACTCAGGCTG AAAAGGTCCC CCAGACCTTT TCACTCTGTA GAGAACTAGG	9180
TGAACTAACA GATGTTTACG AAATTGTCTG GGAAAAGTTG CAGGGTTTAA AATAACAACC	9240
TCAAGGCTGT TTGGGAATAT TGCAAGAGCT GAAATAAAGG AATAAGAATT GATGGATAAA	9300
CGTATTTTGG TTGAAAAAAA GGCTGATTTT CAGGTCAAGT CAGAGAGTTT GGTTAGAGAG	9360
CTCCAGCACA ACTTGGGACT GTCAAGCTTG AAAAGTATTC GTATTGTGCA AGTATATGAT	9420
GTATTTGACT TGGCTGAGGA CTTGTTTGCA CCTGCAGAGA AGCACATTTT CTCTGAGCAG	9480
GTAACCGACC ATGTTTTAGA TGAAGTATCT GTGCAGGCGG ATCTTGCTAA CTATGCTTTC	9540
TTTGCCATTG AAAGTCTGCC AGGGCAGTTT GACCAGCGTG CAGCTTCCTC ACAGGAAGCC	9600
TTGCTTTTGT TGGGAAGTTC GAGTGACGTG ACAGTCAACA CAGCCCAACT TTAAGTGGTG	9660
AATAAGATA TTGATGCGAC TGAGTTGGAA GCTGTCAAAA ACTACCTGCT CAATCCAGTT	9720
GATTCTCGTT TCAAGGATAT CACGACAGGG ATTGCCAAGC AGGAGTTTTT AGAGTCAGAC	9780
AAGACCATTG CCAAAATTGAC TTTCTTTGAA AGCTATGCAG CAGAAGACTT TGCTCGCTAC	9840
AAGGCCGAAC AAGGGATGGC CATGGAAGTG GATGATTTGC TCTTTATCCA AGACTACTTT	9900
AAGTCAATCG GGC GCGTGCC AACTGAGACT GAACTCAAGG TTTTGGACAC TTAAGTGGTCT	9960
GACCACTGCC GTCATACGAC TTTTGAGACA GAGTTGAAAC ACATCGACTT TTCAGCTTCT	10020
AAATTTCAAA AGCAATTGCA GTCAACCTAT GACAAGTATA TTGCCATGCG CGAGGAATTA	10080
GGTCGGTCTG AAAAACCACA AACCTTGATG GATATGGCGA CTATTTTCGG TCCTTATGAG	10140
CGTGCTAATG GACGATTGGA TGATATGGAA GTCTCTGACG AAATCAATGC CTGCTCAGTT	10200
GAAATTGAAG TGGACGTTGA TGGTGTCAAG GAACCTTGGC TCCTCATGTT TAAAAACGAA	10260
ACCCACAACC ATCCAACAGA AATTGAGCCA TTTGGTGGAG CGGCTACCTG TATTGGTGGA	10320
GCTATTCGTG ATCCGTTGTC AGGCCGTTCC TATGTTTACC AAGCCATGCG TATTTAGGT	10380
GCTGGTGATA TTACAGCACC GATTTTCGGAA ACTCGCGCTG GGAAATTGCC ACAACAAGTC	10440
ATTTCTAAAA CAGCAGCTCA TGGTTATTCT TCATATGGTA ACCAGATTGG GCTTGCAACA	10500
ACCTACGTTT GTGAATACTT CCACCCAGGC TTTGTAGCTA AACGTATGGA ACTTGGTGCC	10560
GTTGTTGGTG CCACTCCCAA GGGCAATGTT GTCCGTGAAA AACCTGAAGC AGGTGATGTG	10620
ATCATCCTTC TCGGAGGCAA AACAGGTCGT GATGGTGTCC GTGGTGGAC GGGCTCTTCT	10680

667

AAGGTTCAAA CAGTTGAGTC TGTAGAGACT GCTGGTGCTG AGGTTCAAAA AGGAAATGCC	10740
ATCGAAGAAC GCAAGATTCA GCGCCTCTTC CGTAATGGCA ATGTCACTCG TCTGATCAAG	10800
AAGTCCAATG ACTTTGGGGC AGGCGGTGTC TGTGTGGCTA TCGGTGAATT GGCAGACGGT	10860
CTTGAAATCG ACCTCAACAA GGTGCCTCTT AAATACCAGG GCTTGAATGG TACAGAAATT	10920
GCCATCTCTG AATCACAAGA ACGGATGGCG GTCGTGGTTC GTCCTGAAGA TGTGGATGCC	10980
TTCGTTGCCG AATGTAACAA AGAAAATATT GATGCTGTTG TGGTGGCGAC AGTAACTGAA	11040
AAACCAATC TTGTCATGCA CTGGAATGGT GAGACAATCG TTGACTTGA GCGTCGTTTC	11100
CTTGACACCA ATGGTGTGCG CGTGGTTGTC GATGCCAAAG TTGTGGACAA GGATGTCAAA	11160
CTCCAGAAG AGCGTCAAAC ATCTGCTGAA ACACTGGAAT CAGATACCTT TACGGTTCTA	11220
TCTGACCTCA ACCATGCAAG TCAAAAAGGA TTACAGACTA TCTTTGACTG CTCTGTTGGA	11280
CGCTCAACGG TTAATCACCC ACTTGGTGGT CGTTACCAAC TCACACCAAC TGAGGCATCT	11340
GTGCAGAAAT TGCCAGTTCA ACACGGTGTG ACTCATACTG CGTCGGTCAT TGCTCAAGGT	11400
TTCAACCCAT ATGTAGCTGA ATGGTCTCCA TACCACGGTG CTGCTTATGC GGTATTCGAA	11460
GCAACTGCTC GTTTGGTGGC TGCTGGTGCC AACTGGTTCA AGGCTCGTTT CTCTTACCAA	11520
GAGTATTTTC AGCGTATGGA TAAACAAGCA GAGCGTTTCG GTCAGCCAGT AGCTGCTCTT	11580
CTAGGTTCTA TTGAAGCACA AATTCAGCTT GGCTTGCCAT CTATCGGTGG TAAGGACTCC	11640
ATGTCTGGTA CCTTTGAAGA ATTGACCGTT CCGCCAACCT TGGCTGCCTT TGGGGTGACG	11700
ACGGCAGATA GCCGTAAGGT GCTCTCTCCA GAATTTAAAG CTGTTGGGGA AAATATCTAC	11760
TACATCCCAG GTCAAGCCCT CTCTGCAGAG ATTGATTTTG ACTTGATTAA GAAAAATTTT	11820
GCTCAGTTTG AAGCCATCCA AGCTGACCAT AAAGTGACAT CTGCATCAGC TGTCAAATAC	11880
GGTGGTGTAG TTGAAAGTTT GGCTCTTGCT ACCTTTGGAA ACTATATTGG TGCAGAGGTG	11940
ACCTTGCCCTG AACTTGAAAC AGCTTTGACA GCTCAATTAG GCGGCTTTCT CTTCACATCT	12000
CCTGAAGAAA TTGCTGGAGT AGAGAAGGTT GGACAAACGA AAGCAGACTT TAACTGACT	12060
GTCAACGGTG TGAAGCTAGA TGGACACAAG CTTGACAGTG CATTTCAAGG GACATTGGAA	12120
GAAGTTTACC CAACAGAATT TACCCAAGCG AAAGAAGTAG AAGAAGTACC AGCTGTGGCA	12180
TCAGATGTTG TGATTAAAGC CAAAGAAAAG GTTGAAAAAC CTGTGGTTTA CATCCCAGTC	12240
TTTCCAGGAA CCAACTCAGA ATATGATTCA GCTAAGGCCT TCGAAAAAGA AGGTGCAGAG	12300
GTCAATTTGG TGCCATTCGT GACCTTGAAT GAAGAAGCTA TTGTCAAGTC AGTTGAAACT	12360
ATGGTTGACA ATATCGACAA GACTAATATT CTCTTCTTTG CTGGTGGATT CTCGGCTGCG	12420

668

GATGAACCAG ATGGTTCAGC TAAGTTTATC GTCAATATCC TGCTTAATGA AAAAGTGCGT	12480
GTGGCTATTG ATAGCTTTAT CGCCCGTGGT GGTTCGATTA TCGGTATTTG TAATGGATTG	12540
CAAGCCTTAG TCAAATCGGG TCTCCTACCC TACGGAACT TTGAAGCTGC TAACAGTACT	12600
AGCCCAACCC TCTTCTACAA TGATGCCAAC CAACACGTGG CCAAGATGGT GGAAACTCGC	12660
ATTGCCAATA CCAACTCACC ATGGTTGGTT GGTGTGCAAG TGGGCGATAT CCACGCTATT	12720
CCTGTTTCGC ACGGTGAAGG GAAGTTTGTC GTGACGGCTG AGGAATTTGC AGAGCTCCGT	12780
GACAATGGAC AAATTTTCAG CCAATACGTT GACTTTAAG GTAAACCAAG TATGGATTCT	12840
AAGTACAATC CGAATGGTTC TGTCCATGCC ATCGAAGGAA TTACCAGCAA GAATGGTCAA	12900
ATCATCGGTA AGATGGGCCA CTCAGAACGT TATGAGGATG GTCTTTTCCA AAATATCCCA	12960
GGCAATAAAG ACCAACACCT GTTCGCATCA GCGGTTAAAC ATTTCACTGG AAAATAAGAC	13020
TTACAGATTT TCTAATAGAT AGTATCAGTA ATGTAAAAGT CATGTAAATC TAGCTCTTGA	13080
TGATTACAAA TGAAAATTAG GTATAAAAAA TGACATACGA AGTAAATCT CTTAATGAAG	13140
AATGTGGTGT TTTCCGTATT TGGGGACATC CAGATGCTGC TAAGTTGACC TATTTTGGAC	13200
TCCACAGTCT TCAACACCGT GGTCCAGGAG GGGCAGGAAT CCTCTCCAAT GATCAAGGAC	13260
AACTGAAGCG CCATCGTGAC ATGGGGCTTT TATCAGAAGT TTTCAGAAAT CCAGCTAATT	13320
TGGATAAATT GACAGGAGCT GGTGCGATTG GGCATGTGCG TTATGCGACT GCTGGCGAAG	13380
CTTCTGTAGA TAACATCCAG CCCTTCCTCT TCCGTTTCA CGATATGCAG TTTGTTTGG	13440
CTCATAATGG AAATCTGACC AATGCAGCCT CTCTCAAGAA AGAACTGGAA CAAAGAGGAG	13500
CAATTTTCAG CGCGACTTCG GACTCTGAAA TCTTGCTCA CCTCATTCGT CGCAGTCATA	13560
ATCCTAGCCT GATGGGCAAA ATCAAGGAAG CGCTCAGCCT TGTCAAAGGT GGTTCGCT	13620
ATATCTTGCT GTTTGAGGAC AAGTTGATTG CGGCTCTGA CCCAAATGGA TTCCGACCGC	13680
TTTCGATTGG TAAAATGGCT AATGGAGCAG TTCTGTATC TTCTGAAACC TGTGCTTTTG	13740
AGGTCAATTG TGCCGAGTGG ATTCGTGATT TGAAGCCAGG TGAGATTGTG ATCATTGATG	13800
ACGAGGGCAT TCAGTATGAC AGCTATACAG ATGATACCCA GTTGGCGGTT TGTTCATGG	13860
AGTATATCTA CTTTGCTCGC CCTGATTCTA ATATCCACGG TGTCAATGTC CATACGGCAC	13920
GTAAGAGAAT GGGAGCGCAA TTGGCGCGAG AATTTAAGCA TGAGGCAGAT ATTGTAGTTG	13980
GTGTGCCCAA TTCTTCCCTA AGCGCGGCTA TGGGATTGTC GGAAGAATCA GGCTTACCAA	14040
ATGAAATGGG TCTGATCAAA AACCAATACA CCCAGCGAAC TTTTATCCAA CCGACTCAAG	14100
AATTGCGGGA GCAAGGAGTG CGGATGAAAC TGCTGCTGT TTCGGGTGTT GTCAAAGGCA	14160
AACGTGTGGT CATGGTGGAT GATTCCATTG TACGTGGAAC AACCTCTCGT CGTATCGTTC	14220

669

AGCTCTTGAA AGAAGCGGGT GCGACTGAGG TTCACGTTGC CATTGGAAGT CCTGCACTAG	14280
CGTATCCATG TTTCTACGGG ATTGATATCC AGACCCGTCA GGAGCTGATT GCAGCCAATC	14340
ATACGGTCGA AGAAACTCGC CAAATCATTG GTGCGGACAG TCTGACTTAT CTTTCAATTG	14400
ATGGCTTGAT TGAGTCGATT GGTATCGAAA CAGATGCGCC GAACGGTGGT CTCTGTGTCG	14460
CTTACTTTGA CGGTGACTAC CCAACGCCTC TTTATGACTA CGAAGAAGAC TATCGTAGAA	14520
GTTTGGAAGA AAAGACCAGT TTTTACAAGT AGGCGACAGA TTCTCCATTA AAGAAAAGGA	14580
AAAAATAAAT GACAAATAAA AATGCATATG CCTCACGTCT CACTACTGAC TAAAGGCTTA	14640
AGCATTTAGT CAGTAGACGC TTTGTCTTAT AGGATCAAAG CTAGAGCCCT GACTAGTATT	14700
TTTAGATAAA AAGATGGTTT ATCTAAAAAT ACGTCGCAGT CTTTCTCAA AAAAGAAAAG	14760
GAAAAATAAA ATGGCAAATA AAAATGCGTA CGCTCAATCT GGTGTGGATG TTGAAGCGGG	14820
TTATGAAGTT GTTGAACGGA TTA AAAAGCA CGTGGCCCGT ACGGAGCGTG CAGGTGTCAT	14880
GGGAGCTCTT GGTGGCTTTG GTGGTATGTT TGACCTTTCC AAGACTGGGG TTAAGAACC	14940
CGTCTTGATT TCAGGGAAGT ACGGTGTCGG AACCAAGCTC ATGTTGGCTA TCAAGTACGA	15000
CAAGCACGAT ACCATCGGGC AGGACTGTGT GGCCATGTGT GTCAACGACA TCATTGCTGC	15060
AGGTGCGGAA CCCCTCTATT TTCTCGACTA CGTAGCGACA GGGAAGAATG AACCAGCTAA	15120
GCTAGAACAA GTGGTTGCTG GTGTGGCAGA AGGTGTGTGT CAGGCTGGTG CTGCCCTCAT	15180
CGGTGGGGAA ACGGCTGAAA TGCCGGGCAT GTACGGCGAA GACGACTATG ACTTGGCTGG	15240
TTTTGCGGTC GGTGTGGCTG AAAAATCTCA AATCATTGAC GGTTCAAAGG TGGTAGAGGG	15300
AGATGTTCTT CTCGGAAGTG CTTCAAGTGG GATTCACTCA AATGGTTACT CTTGGTTCTG	15360
TCGTGTCTTT GCGGATTACA CAGGTGAGGA AGTCCTACCA GAATTGGAAG GCAAGAACT	15420
TAAGGAAGTT CTAAGTGAGC CGACTCGTAT CTATGTCAAG GCTGTCTTGC CGCTCATCAA	15480
AGAAGAGTTG GTCAACGGCA TTGCCACAT CACAGGTGGT GGCTTTATCG AAAATGTCCC	15540
TCGTATGTTT GCAGATGACC TAGCTGCTGA AATTGATGAA AGTAAAGTTC CAGTGCTTCC	15600
AATTTTCAA ACCCTTGAAA AATACGGTCA GATTAAACAC GAAGAAATGT TTGAAATCTT	15660
CAATATGGGT GTGGGACTTA TGTTGGCGGT CAGCCCTGAA AATGTAGAGC GTGTAAAAGA	15720
ATTGTTGGAT GAAGCAGTCT ATGAAATTGG TCGCATCGTC AAGAAAGAAA ACGAAAGTGT	15780
CATTATCAA TGAAAAAAT AGCGGTTTTT GCCTCTGGTA ATGGCTCAA TTTTCAGGTG	15840
ATTGCCGAAG AATTTCCAGT GGAGTTTCTC TTTTCAGACC ATCGTGATGC CTATGTGCTT	15900
GAGCGTGCAA AGCAGCTCGG CGTTCTGTCC TATGCTTTTG AACTCAAGGA GTTTGAGAGC	15960

670

AAGGCAGACT	ACGAAGCAGC	CCTTGTCGAA	CTCTTGGAAG	AACACCAGAT	TGACTTGGTT	16020
TGCCTAGCAG	GCTACATGAA	AATCGTTGGA	CCAACCTTAT	TGTCGGCTTA	TGAAGGTCGG	16080
ATTGTCAACA	TTCATCCAGC	CTACTTGCCA	GAATTTCCAG	GAGCTCATGG	GATTGAGGAT	16140
GCTTGGAATG	CTGGCGTGGG	TCAGTCTGGT	GTGACCATTC	ACTGGGTGGA	TTGGGTGTG	16200
GATACAGGCC	AGGTCATCAA	ACAGGTTCTG	GTGCCACGAC	TAGCTGATGA	TACCATTGAC	16260
AGATTTGAAG	CTCGCATCCA	TGAAGCAGAG	TACAGGCTGT	ATCCGGAAGT	AGTGAAGGCT	16320
CTATTTACAG	ATTGACTTTT	TGATGATTCA	TATGATATCT	TTGATTTTAA	ATTGGAGTCA	16380
GTGTTTGTG	AAGACGGCTT	CAAACGGAGG	TATTTGTAAT	GTTAGAATCT	AAAAAAACAA	16440
CTCGATATGT	ATTTTATGTC	TATCTGATGT	TATTAACCTG	GGGAATCTTA	TTTAAGTTTG	16500
AAACAAATCC	TGAATTTATA	GCATTTTCT	TAGCTCCAAG	GTATATCAAT	TGGATTCCAT	16560
TTTCAGAACC	ACTAATAGTC	GATGAAAAA	TTGTTTTTGC	TGAAATGTTA	TTTAATCTGA	16620
TTTTCTTTAT	TCCATTAGGT	GTTTGTTC	CTTTGATAAA	AACTAATTA	TCTAGTTTAA	16680
GAATAGTCGG	GACAGGTTTC	TTGATTAGTT	TATTGTTTGA	GTGCTTACAG	TATATTTTAG	16740
CAATAGGTAT	AACAGATATA	ACGGATTGTA	CTTTAAATAC	GCTAGGTGTC	TGTGTAGGCT	16800
TACTGATTTA	TCAAATTTTT	ATAAGAGTGT	TCAAATCACA	GACTAGAAAA	TGGATCAATA	16860
TCTTAGGTAT	GCTTAGCCTT	GGTTTTGCTT	ATCTTGTTTT	ACTGTTACTG	CATTTACTTA	16920
GTGTTTAACT	AATGATTAAA	AAGGAGAATA	TAATGACTAA	ACGCGTCTTA	ATCAGCGTCT	16980
CAGACAAAGC	GGGCATTGTT	GAATTTGCCC	AAGAACTCAA	AAAACCTGGT	TGGGAGATTA	17040
TCTCAACAGG	TGGAACCTAAG	GTTGCCCTTG	ATAATGCTGG	GGTGGATACC	ATTGCTATCG	17100
ATGATGTGAC	TGGTTTCCCA	GAAATGATGG	ACGGTCGTGT	GAAGACCCTC	CACCCAAATA	17160
TCCACGGAGG	GCTTCTCGCT	CGTCGTGACT	TGGATAGCCA	CTTGAAGCG	GCTAAGGACA	17220
ACAAGATTGA	GCTCATTGAC	CTTGTTGGTG	TCAACCTTTA	CCCATTTAAG	GAAACTATCC	17280
TTAAACCAGA	TGTGACTTAT	GCTGATGCAG	TTGAAAATAT	CGATATTGGT	GGGCCATCTA	17340
TGCTTCGTTT	AGCAGCGAAA	AATCATGCCA	GTGTTACAGT	TGTGGTAGAT	CCTGCTGACT	17400
ACGCTGTGGT	TTTGATGAA	TTGGCAGCAA	ACGGCGAAAC	CTCTTATGAA	ACTCGCCAAC	17460
GTTTAGCAGC	CAAAGTATTT	CGTCACACAG	CGGCTTATGA	CGCCTTGATT	GCAGAATACT	17520
TCACAGCTCA	AGTGGGTGAA	AGCAAGCCTG	AAAAACTCAC	TTTGACTTAT	GACCTCAAGC	17580
AACCAATGCG	TTACGGTGAG	AATCCTCAAC	AAGACGCGGA	CTTTTACCAG	AAAGCTTTGC	17640
CTACAGACTA	CTCCATTGCT	TCAGCCAAAC	AGCTCAACGG	GAAAGAATTG	TCATTTAATA	17700
ATATCCGTGA	TGCAGATGCT	GCTATCCGTA	TCATCCGTGA	CTTCAAAGAT	AGTCCAACCG	17760

671

TTGTGGCTCT CAAACACATG AATCCATGTG GAATTGGTCA AGCTGATGAC ATCGAGACTG	17820
CTTGGGACTA CGCTTATGAG TCTGACCCAG TGTCTATCTT TGGTGGGATT GTCGTCCTCA	17880
ACCGTGAGGT GGATGCTGCG ACAGCTGAGA AGATGCACGG CGTTTTCTCT GAAATCATCA	17940
TTGCACCAAG CTATACGGAT GAAGCGCTAG CCATTTTGAT CAATAAAAAG AAAAATTGTC	18000
GTATCCTTGC CTTGCCATTT AATGCTCAAG AGGCTAGCGA AGTGGAAGCA GAATACACAG	18060
GTGTAGTCGG TGGACTTCTC GTGCAAAATC AAGACGTGGT CAAGGAAAGC CCAGCTGACT	18120
GGCAAGTGGT GACTAAACGT CAGCCAACGT AGACAGAAGC GACTGCTCTT GAGTTCGCTT	18180
GGAAGGCTAT CAAGTACGTC AAATCAAATG GTATTATCGT GACCAACGAC CACATGACAC	18240
TTGGTGTGG TCCAGGTCAA ACCAACCGTG TGGCTTCTGT TCGCCTTGGC ATTGACCAAG	18300
CCTAAGATCG TCTGGACGGG GCGGTCCTTG CTTGAGATGC CTTCTTCCCA TTTGCGGATA	18360
ACGTGGAAGA AATCGCCAAA GCAGGAATTA AGGCCATCAT CCAGCCCGGT GGCTCTGTCC	18420
GTGACCAAGA ATCCATCGAA GCAGCGGATA AATACGGCTT GACTATGGTC TTTACAGGTG	18480
TGAGACATTT TAGACATTAA GAAGATAAAA GGGAAGAAAA CAGTTTCTTT CCTTTTTTGG	18540
CTTAAATAC TAACTGAAAC AAGATTAAAA CGAATTTTTT TGATATAATG TTGTTAAATA	18600
ATTCGCAAAA GAGGTTGAGG AATGAAACTG CTTGTTGTCG GTTCTGGTGG TCGTGAGCAT	18660
CGGATTGCTA AAAAGTTACT TGAATCAAAA GACGTGAAA AAGTCTTTGT AGCTCCTGGG	18720
AATGATGGGA TGAATCTGGA TGGTTTGGA TGGTTAAATA TCTCTATTTC CGAACATTAT	18780
AAATTGATTG ACTTCGCAAA GACCAATGAT GTTGCTTGA CCTTTATCGG TCCAGATGAT	18840
GCCCTTGCTG CTGGTATCGT GGATGATTTT AACCAAGCTG GACTTAAGGC CTTTGGTCCG	18900
ACTAGGGCTG CAGCGGAGCT GGAGTGGTCC AAGGATTTCT CCAAGGAAAT CATGCTCAAA	18960
TACGGCGTTC CGACAGCAAC ATATGGCACA TTTTCAGATT TCGAGGAAGC CAAAGCCTAT	19020
ATCGAAAAGC ATGGTGCGCC TATCGTAGTC AAGGCGGATG GCTTGGCACT TGGGAAGGGT	19080
GTCGTCGTTG CTCGACGGT TGAGCAAGCG GTCGAAGCCG CTCATGAGAT GCTTTTGGAC	19140
AATAAATTG GTGACTCAGG TCGCGCGCTG GTTATTGAGG AATTCCTTGA AGGAGAGGAA	19200
TTTCACTCT TTGCCTTTGT CAATGGTGAT AAGTTCTACA TCATGCCAAC GGCTCAGGAC	19260
CACAAACGTG CCTATGATGG CGACAAAGGG CCTAACACGG GTGGTATGGG TGCCTATGCG	19320
CCAGTCCAC ACTTACCACA GAGTGTAGTT GATACAGCGG TTGACACCAT TGTCAAGCCA	19380
GTTCTAGAAG GGGTGATTAA AGAAGGTCGC CCTTATCTGG GAGTTCTTTA CGCAGGGCTT	19440
ATCCTGACAG CTGATGGACC GAAAGTCATT GAGTTCAACG CTCGGTTCGG AGATCCAGAA	19500

672

ACTCAGATTA TCTTGCCCTCG CTTGACCTCT GACTTTGCTC AAAATATCAC AGATATCCTG	19560
GATAGCAAGG AGCCAAATAT CATGTGGACG GACAAGGGTG TGA CTCTGGG TGTGGTTGTC	19620
GCATCCAAGG GCTACCCGCT AGACTATGAA AGGGGCGTTG ACTTGCCAGC CAAGACAGAA	19680
GGCGATGTCA TCACCTACTA TGCAGGGGCT AAGTTTGCGG AAAATAGCAG AGCACTGCTC	19740
TCAAACGGCG GACGAGTTTA TATGCTCGTT ACCACAGCAG ATACCGTCAA AGAAGCCCAA	19800
GCCAGCATAT ACCAAGA ACT ATACCAACAA AAAATAGAAG GACTCTTCTA CCGAACAGAT	19860
ATCGGAAGCA AGGCAATTAA GTAAAGATAT AAGAATAACG CGCCGTAGTC GCCAAACACG	19920
ATAATGGTCG TCGTGGTGAA AAGACCAGAA CAGTGAATGT TCTGGTCAGG GGGAACTTG	19980
GAGACCTTAG GCTCAAAGTT TAGGAATGAA ACCGAAGGTT TGCTTCCGCC TCCATCACCT	20040
AAGACCATTA TCAAAAAGAA AAATAAAAAT TCACAAAATA CGTTAATGAT CGTATGGTTT	20100
GCGAGCGTTA GCGAGCTAAT ATAGAACAAT CACCGCGGTT GTGAAAGAAC GATTGGATGA	20160
TAATCCAATC GTTCAGGGAA ATTGGAAGAC CTTGGGTTTC CAATTTAGGC ATGAGACACC	20220
TTTGGTGGCT GCTGCCGTCC CTCACAAGCT AAGGTGATTG TTGAAAAAGA GGAAAAAGGA	20280
GAAGAAATGA AACCAGTAAT TTCCATCATC ATGGGCTCAA AATCCGACTG GGCAACCATG	20340
CAAAAAACAG CAGAAGTCCT AGACCGCTTC GGTGTAGCCT ACGAAAAGAA AGTTGTTTCC	20400
GCACACCGTA CACCAGACCT CATGTTCAA CATGCAGAAG AAGCCCCGTAG TCGTGGCATC	20460
AAGATCATCA TCGCAGGTGC TGGTGGCGCA GCGCATTGTC CAGGCATGGT AGCTGCCAAA	20520
ACAACCCTTC CAGTCATTGG TGTGCCAGTC AAGTCTCGTG CTCTTAGTGG AGTGGATTCA	20580
CTCTATTCTA TCGTTCAGAT GCCGGGTGGG GTGCCTGTTG CGACCATGGC TATCGGTGAA	20640
GCTGGAGCGA CTAACGCAGC TCTCTTTGCC CTCGCTCTCC TCTCTGTAGA AGATAAGTCC	20700
ATTGCGGATG CACTTGCCAA CTTTGCTGAA GAACAAGGAA AAATCGCAGA GGAGTCGTCA	20760
AATGAGCTCA TCTAAAACAA TCGGAATTAT CGGTGGCGGT CAACTGGGTC AGATGATGGC	20820
CATTCTGCT ATCTACATGG GCCACAAGGT TATCGCGCTG GATCCTGCGG CGGATTGCCC	20880
GGTCTCTCGT GTGGCGGAAA TCATTGTGGC ACCTTATAAC GATGTAGACG CCCTCCGTCA	20940
GTTGGCAGAC CGTTGCGATG TCCTCACTTA TGAGTTTGAA AATGTCGACG CTGACGGTTT	21000
GGATGCCGTT ATCAAGGATG GACAACTCCC TCAAGGAACA GATCTGCTCC GCATTTGCGA	21060
AAATCGTATT TTTGAAAAGG ACTTTTGTG AAACAAGGCT CAAGTCACTG TGGCACCTA	21120
CAAGTCTGTG ACTTCTAGCC TAGACTTGGC AGATATCGAC TTGTCGAAAA ACTATGTCCT	21180
CAAGACTGCG ACTGGTGGCT ACGATGGTCA TGGACAAAAG GTTATTCGTT CAGAAGCAGA	21240
CTTGGAAGCA GCCTATGCGC TAGCAGACTC AGCAGACTGC GTCTTGAAG AATTTGTCAA	21300



673

CTTTGACCTT GAGATTTCTG TCATCGTGTC AGGAAATGGC AAGGAGGTGA CGTTTTTCCC	21360
AGTTCAGGAA AATATCCACC GCAACAATAT CCTGTCTAAG ACCATCGTAC CAGCCCGCAT	21420
TTCTGAAAAGT CTAGTAGACA AGGCTAAAGC TATGGCAGTG CGAATCGCAG AACAACTCAA	21480
CTTGTCTGGA ACTCTCTGTG TGGAAATGTT TGGCAGAGCT GATGACATCA TTGTCAATGA	21540
AATCGCCCCA CGACCACATA ACTCTGGGCA CTATTCTATT GAAGCCTGTG ATTTCTCTCA	21600
GTTTGACACC CATATTCTGG GTGTTCTCGG AGCACCATTA CCAGTCATCA AACTCCATGC	21660
GCCAGCCGTT ATGCTTAAATG TCCTCGGTCA GCATGTCGAG GCTGCTGAAA AATATGTCAC	21720
AGAAAATCCA AGCGCCCACC TCCACATGTA TGGTAAAATA GAAGCAAAGC ATAATCGTAA	21780
GATGGGACAT GTGACTTTGT TTAGTGATGT GCCGGATAGT GTGGAAGAGT TTGGGGAAGG	21840
GATTGATTTT TAGGACAAGT CTATGATACA AATTATCGTT AATACATTTA TTGAAAAGTA	21900
TAAGACTGGA GCAGTTGTTG AAGTGTTGTA TGCCAGTGCT GACCAAGATA AGGTACAAGC	21960
TAAATATGAA GAACTAGCTG CACAATACCC CGAAAATTAT TTAGCTATCT ATAATGTACC	22020
GCTGGATACG GATTTGAATA CACTAGATCA TTACCCGCTCT GTGTTTATTG GAAAAGAGGA	22080
GTTTGAGTAG AAATCTTGGT TTACCTAGAT AGCTTATTCC CAACAGCTTA AGAAGAAAGG	22140
AAAAATTAAC ACATGATCAA CCGTTACTCT CGCCCTGAGA TGGCGAATAT TTGGAGTGAA	22200
GAAAATAAAT ACCGTGCTTG GCTTGAGGTG GAAATCCTCT CTGACGAGGC ATGGGCTGAG	22260
TTGGGGGAAA TCCCTAAGGA AGATGTGGCT TTGATTGCGA AGAAGGCGGA CTTTGACATC	22320
GACCGTATTT TGGAAATTGA GCAGGAGACG CGCCACGATG TGGTGGCTTT CACGCGTGCG	22380
GTTTCTGAGA CTCTTGGTGA AGAGCGCAAG TGGGTTCACT ATGGGTAAAC TTCTACTGAC	22440
GTGGTGGATA CTGCTTATGG TTACCTCTAC AAGCAGGCCA ACGACATCAT CCGTCGTGAC	22500
CTTGAAAAC TCACTAATAT CATCGCTGAC AAGGCCAAGG AGCACAAGTT CACCATCATG	22560
ATGGGGCGTA CTCATGGTGT GCACGCTGAG CCGACAACCT TTGGTCTTAA ATTAGCAACT	22620
TGGTACAGCG AAATGAAACG CAATATCGAG CGCTTCGAGC ATGCGGCTGC TGGTGTAGAA	22680
GCTGGTAAGA TTTCTGGTGC GGTGGGAAC TTTGCCAATA TCCCACCATT TGTAGAGGAG	22740
TATGTCTGCG ATAAACTTGG CATCCGTGCC CAAGAAATCT CTACACAAGT CCTTCTCGT	22800
GACCTTCACG CTGAGTACTT TGGGTTCTT GCCAGCATTG CGACTTCAAT CGAACGTATG	22860
GCGACTGAGA TTCGTGGTCT ACAAAAATCT GAGCAACGCG AAGTAGAAGA GTTCTTTGCT	22920
AAAGGGCAAA AAGGCTCTT AGCAATGCCT CACAAACGCA ACCCAATCGG TTCTGAAAAT	22980
ATGACTGGTC TGGCGCGTGT CATTCGTGGT CACATGATTA CGGCTTATGA AAACGTCGCT	23040

674

CTCTGGCATG AACGCGATAT TTCTCACTCA TCAGCTGAGC GTATCATCAC ACCAGATACG	23100
ACCATTTTGA TTGACTACAT GCTCAACCGT TTTGGAAATA TCGTCAAGAA CTTGACAGTC	23160
TTCCCAGAAA ATATGATCCG AAACATGAAC TCGACTTTTG GTCTTATCTT TAGCCAACGG	23220
GCTATGTTGA CATTGATTGA AAAAGGCATG ACCCGTGAGC AAGCCTATGA CTTGGTGCAA	23280
CAAAAACAGC CTACTCTTGG GACAACCAAG TAGACTTTAA ACCACTTCTT GAGGCAGATT	23340
CAGAAGTAAC ATCACGTCTC ACACAAGAAG AAATCGATGA AATCTTCAAC CCAGTTTATT	23400
ACACCAAACG AGTGGATGAT ATCTTTGAAC GTCTTGACT AGGTGATTAA TAAAAAATA	23460
AACAGCGAGC TTCAATCTCG CTGTTTATTT TTTATCGAAA AGACTTAGTC TTCTTTTCTT	23520
TTAGTGAGTC CATAGGCTGC TAGTGTGGAC ATGAGTCCTG CGACTACTAG TCCTGCAGAA	23580
TCGTGAGTTC CTGTTTCAGG AAGTTTTTTC TCTGTTACCA CAGGAGCTGG ATCTTGAGGA	23640
AGAACTTTGC TTTCTCAGC AGGAGCAGTT GATGGAGCTG GTTGGCTTGG GATTTCTAGT	23700
TTTGGTTTTT CTTCAGCAAT AGCGGCTTGT CCGTTTTCAT CGCCTACATG TGTTACCATA	23760
GTTCCGACTT CGACTATTTG AGTAACGGCT TCCTGTGCTA CGACACTATT TACAAGTGTT	23820
TTCACTTCCT TACCATCGGC AGAAGTGCTC ACAGAGTAGA AGTTGCTACG ATGTCCATTG	23880
ACGCCCTTAG TAATGACTTG TGTTTTTCCT TTGAGTAAGA GTGGATTTTC ACAAGTCACT	23940
GTGGTAAATG GAATTTCTTC TTCTTGATA TCCAGTCTAG GTTTTACCTC AGTAGTTGGT	24000
GCAAGACCAC TTTCATCACC CTTGTGAGTT ACAGGAGCGC CAACTTCAAC CACTTGTTTT	24060
ATAACTTCTT TGGTTACCTG GCTATCAAGG ACTGTTTCTG TTGTTTTTCC ATTTTCAGTG	24120
AGTACAGAGA TGTAAAGAGT TCGTTCACCT TTGACTCCTG CTGTGATAAT ATTTTCCTGA	24180
CCGGCTGGGA GGTTAGGATT TTCTTTCTTG ATAACCTCAA ATGGAATTC TTCAGTTCTT	24240
GTGATGAGTT CTGGTCTGGT TTCAACATTG GCAGCCACTT CATTTTCATC TAGGCTTCCT	24300
GAATGAGTTA CAGCTGGTTT GAGGCCTTGA AGAGCGGCTT TTAGGTTGGC TACAAGCGTG	24360
TCAAGCTCAG CTTGTTTATT ACGGTTGAGG TTGTAATTGA GAGCTGTTTT AGCTGCGTCA	24420
AGGGCCTCAA GACTTTCTTT ACTATATCCT TCTAAGTTTG TAGGAATTTT AGCTAATTCT	24480
TCGCGGAGAG CATTATAATT AGCACGAAAG TAGTCTTTGT TGTGGTCTGC AAAGGCAGTC	24540
ATGAGTTCAA AGATTTCTCT TTCTTGTAT TCAGCGCTTG GTCTATCTGC CCAGATTGAA	24600
AGCATACTTC CGACTGTTGG AAGATCTACT TCAGGATATT TGGTAGAAGC TAGTTGATTG	24660
AATGGTGTTT TTCCAGTATT CTCAATAGCT TTCTTGAGGA AACCACCACC ATCTTCTGGT	24720
TTTTGACCAA GAATGTAGTA CCAGTCACCG TTGGTATTCA AGAATTTATA GCCTTTGCTT	24780
GCTAGGTATT GAGGTGATGC GAGGTTATAT CCCCACCAGC CTTTAGACCA GTAAGAAATC	24840

675

AAGACATCTT TGTCAAACG AACATCGTCC TTGTCTTCAT AGTAGAAGCC ATCGTTGAAG	24900
GCCATTGGTT GAAGCCCTCT TTCTTTGGCC ATAGCTGCGA GGGTGTGGC ATATTCGGCA	24960
AATTTGCCAT AGAGTTGATA CCACTTGAGG TAGTACCAGC CTTGGGCACT AGTCGCATCG	25020
TTGGCGTATT CGTCAGTACC AAAGTTGAAA ATCTTTGTTT TACCTGCAAA GAAGTCCATG	25080
TATTTACCGA TGAGGGCTTT TACAAAGTTC ATCGCTTCTT CGTTTTTCAA GTCCATAGTT	25140
GTTTTTGAAA CTTTATCAAA GTGGGCTTGA GGATTTTAA TACCTAATTT TTCCATGGCA	25200
ACCAGCATAG CATCCATGTG ACCTGGACTG TTAATAGCTG GGATGAGACC GATGTCCCTA	25260
GATTTAGCGT ATTCAATTAG CTCTGTTACT TCTGCCTGTG TTAGTGCACT ACCGTTTGGA	25320
TCGTCGTAGT AAGCTTTAGT TCCTTCGATA ATAGCTTTTT TAACGTCATC ACTAGCATAG	25380
GTTTTTCCGT TGGCAGTAAT GGTCAATCA TCGAGTAGAA AGCGAAGTCC GTCATTTCCCT	25440
AGAAGGAGAT GGACATCAGA ATATCCGAGC TCACTGGCCT TGTCTACGAT GCGTTTGAGC	25500
TGGTTCAGAG TAAAGTATTT GCGTCCAGCA TCGATTGAGA TTACCTTGTT TTTGGCAAGT	25560
TTTTCAACCT CACGTTTAGC TTCTTCTTCT TTTTGAGCTT CAGGCGTGAG GGTCAAGTTG	25620
TTGACAGTTT CTTGAAGTTT AGCAATGGCT TGATCAATCG TATCTTGTTG GGCACGGCTA	25680
AGGTTGCTAT CGAGAGAGCG AATAGCTTTT TCAGCTTCTT TTACGGCCGT GACGCTTCT	25740
GCAGTATAAC GGTTCAGGTC TTTTGGTACC TCGTTAAGTG CTTGCTCTGC AGATTCTATA	25800
TCAGCTGCGA AGTATTCAGC GTTGGCATT TGCAAAATGAC GCATGAGTTT GAAGAGGCGT	25860
GATGGTGAAT AACGTGCAGA TGGAGTGTC GCCCAAGCAG CTACCATAACC ACCGATGATT	25920
GGGATATCAG CTCCTTCTGT TTTTGGTACA GAAGTGATTG GTGTGTTTTT AATACCATTG	25980
AGCCCCGAT CGAGATTGTA CCAGCCTTGG CCATCAGCGT TTCGTCCAAG AACGTAGTAC	26040
CAAGCATCAT TGGTATTAAG GATTTGGTGA CCTTTTTTCAG CTAGTAGTTT AGAAGAAGCG	26100
ACATCGTAGC CTCCCCAACC ACCAGTCCAC ATAGAAACGA TGATGTCTTT GTCAAAACTA	26160
CCAAAGCTTG TGTGCTATT GTAGTAGATA CCGTCGTAA AAGCCATTGG TTTGAGACCG	26220
TGCGATTTTA CAATACGAGC GAGGTCATTG GCGTAGGCAA TAAATTTTTC ATAGCCTTTT	26280
ACAGGGTAGC CTTGCTTTGG ATAGATTTTA TCAGCTTGAA GCACACTCCA ACCTTTAGCA	26340
TCTGTCGCAT CATTGGCATA TTCATCAAGT CCGATGTTGA AGATTTCACT CTTTTCCCG	26400
AAATAAGCAG CATACTTGTC GATAAGGGCT TTTGTAAAAG CGACAGCTTG TTCGTTGTCA	26460
AGATCGACAG TACGGGCTGA TTTCTTCCCA AAATAGCTAA AGTTAGGGTT TTGGATTCCC	26520
AATCTTTTCA TGGCATTGAG AATCGCATCC ATGTGTCCAG GACTATTTAC TGTCGGAATG	26580

676

AGACCGATAC CTTTATCTT GGCATAGTTA ATCAGATCTG TCATTTGACT TTCTGTTAAG	26640
TGATTGCCGT TTGGATCGTT GTAATAATCA TTTGTACCTT TTTCAATGGC GCGTTTGACA	26700
TCGTCACTGG CATAGGTCTT GCCGTTAGCT GTGATGCTCA TATCGTCCAA CATGAAACGG	26760
AGTCCATCAT TTCCGACTAA TAGTGTAATA TCAGTGTAGC CATAATGTTT CGCTTTATCG	26820
ATGATTTCTT TGAGCTGTTT TGGTGAGAAA TATTTACGTC CAGCATCAAT AGAAACAATT	26880
TTCTTTTTTC CTAGTTTTTC ATTTACAGTT GCAGCACGTT CCTTTCCTGC CTCTGTTGCC	26940
GGTTTGTGAG CCTCTGCTTT CGCTTCATCT TTTTGTAGCT GTTTATCCTT GTCAGTCTTG	27000
TCTGTATTG ACTCTTTAGA ATCAACCTCT TTCGCTTCTT CCTTTTTAGG GCTAGCTTCT	27060
TCTGCCTTTT TATTAGCAGT TTCTTTTTCA GCAGAAGTTG GAGTTACCAC TTCTGCTTTA	27120
TCACTAGGAG TTGAACAAAC TTCCTCTTGT GGTTTTTCTT CTGTTTTTGG AAGACTAGCT	27180
ACCTTATCAG TAGCTGGAGT TTCTGTTTCT ACAGTTTTTG GAGCTTCTGG TTGAAGCACT	27240
GCTTTAGGTG TTTCTCAGT CCGATTTTCG GATGATTGAG GGAATCAGA AACCGTATGG	27300
ATGGTCGGT GGTTTTCTGT AGTAGTAGGA GTAACCCAT CGGCTGCAAC AGTCTGTGCT	27360
TGGAAGGCAA ATCCAATTAG AACAGAAGCT GCTCTACAG CGTATTTACG AATAGAAAAA	27420
CGCTGTTGTT TTTTATGTTT CATTGCAAAA CCTCCTGATT GCATTGTTAT ATTGATAGCG	27480
ATTATATAAA TCAACGCCTT TATTTTATTT CTTATATTAA TTTCTTATAT TAACGAGAGT	27540
CAAGAGGAGA TGACAAAAAA CTATAATAAG TATAAAAAAA TATAAAATTT AAACCTAAGA	27600
TTTCAGATTG GTCGGAAAAA ATACGTATAT ATATCTAGTA TAATTTTTGG TTCTATTTCT	27660
ATAAAATATT CCACAAATTA TAGAATTTTC CAAAAATAGG TAAGCGCTAC CTTTTTGGTG	27720
TAGTATAATA AGCATAGAAA AAGCCCAAGC GATTAGCTCA GGTTTTCTTC TTAGTGATCA	27780
CGGTACATG AGATAAATTT AATCTTGTAG TAATCAGATC GTTTGTAAGT TTCACTGTAT	27840
TCTAAACTT GGCCAGTTGA TTCGAGTTTG GTGATTTTAG TTTGTAGGAC AGTAGGGAAT	27900
TGTTTCATCGA CTCCGAGGAC TGAAGCTGCA TGTTCTGGAG TTGGAAAGAC TATTTCTGTTG	27960
ATTTCTTCAA AGTGTTTCATC ATTCATGTGA ATGTGGTAGT CTAACCTGAA ACGATTATAG	28020
ATAGAATAT AGTATTCAAG GTTTGGATAA TTTGCGTTGA TATATTGTTT TGGGATGTAG	28080
GATGTATGGT AGATATAAAC GACACCGTTT GATTCGCGGA TACGTTCAAT CTTGTAGTAG	28140
AATTGATCGC CGCGTAGACC CAATTTTTCC AAGTAAACAA GCTTGTTCCT CCGTTCAATT	28200
GAAAGAACAG TTACCTTATC ATCTTTAGCA TTGAAGAGTT CAATATCTGA AAACCTTACA	28260
AGCTTGTGTT TGCGTGACG TGAAACGAAG GTTCCTTTTC CTTGTTGGCG GACAATATAG	28320
CCATCTTTGG CAAGGTCGTT TAAGGCGCGA ACAACTGTGA TAGAGCTGAC ATCGTACATT	28380

677

GAAATGAGTT CTGCTTCAGT GTAAAATTTA TCTCCACTGC TAAACTGCCC AGAGATGATT 28440

TTATTTTSTA ATTCTCTTT TATGTATTGA TGG 28473

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6749 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

CCTGATGGGT GGTATGCGAG GATACAGTTC TGAAATCCG CGTTACTTAA TTAATGGACG	60
CGAAGTCACA CCTGAGGAAT TTGCTCACTA TCGTGCGACT GGTCAATTAC CAGGAAATGC	120
AGAACTGAT GTGCAATGC CACAACAGGC ATCAGGTATG AAACAAGGCG GTGTCTTGC	180
AAAAC TAGGT CGAACTTAA CAGCAGAAGC GCGTGAGGGC AAGTTGGATC CTGTPATCGG	240
ACGAAACAAG GAAATTCAAG AAACATCTGA AATCCTCTCA CGCCGCACCA AGAACAATCC	300
TGTTTTGGTC GGAGATGCAG GTGTTGGTAA GACAGCAGTT GTCGAAGGTC TAGCGCAAGC	360
CATTGTGAAC GGAGATGTTT CTGCTGCTAT CAAGAACAAG GAAATTATTT CTATTGATAT	420
CTCAGGTCTT GAGGCTGGTA CTCAATACCG TGGTAGCTTT GAAGAAAATG TCCAAAACCT	480
AGTCAATGAA GTGAAAGAAG CAGGGAATAT TATCCTCTTC TTTGATGAAA TTCACCAAAT	540
TCTTGGTGCT GGTAGCACTG GTGGAGACAG TGGTTCTAAA GCACTTGCGG ATATTCTCAA	600
GCCAGCTCTC TCTCGTGAG AATTGACAGT GATTGGGGCA ACAACTCAAG ACGAATACCG	660
TAACACCATC TTGAAGAATG CTGCTCTTGC TCGTCGTTTC AACGAAGTGA AGGTCAATGC	720
TCCTTCGGCA GAGAATACTT TAAAATTCT TCAAGGAAT CGTGACCTCT ATCAACAACA	780
CCACAATGTC ATCTTGCCAG ACGAAGTCTT GAAAGCAGCG GTGGATTATT CTGTTCAATA	840
CATTCTCAA CGTAGCTTGC CAGATAAGGC TATTGACCTT GTCGATGTAA CGGCTGCTCA	900
CTTGCGGCT CAACATCCAG TAACAGATGT GCATGCTGTT GAACGAGAAA TCGAAACGGA	960
AAAAGACAAG CAAGAAAAAG CAGTTGAAGC AGAAGATTTT GAAGCAGCTC TAACTATAA	1020
AACACGCATT GCAGAATTGG AAAGGAAAAT CGAAAACCAC ACAGAAGATA TGAAAGTGAC	1080
TGCAAGTGTC AACGATGTGG CTGAATCTGT GGAACGAATG ACAGGTATCC CAGTATCGCA	1140
AATGGAAGCT TCAGATATCG AACGTTTGAA AGATATGGCT CATCGCTTGC AAGACAAGGT	1200
GATTGGTCAA GATAAGGCCG TAGAAGTTGT AGCTCGTGCT ATCCGTCGTA ACCGTGCTGG	1260

678

TTTTGATGAA GGAAATCGCC CAATCGGCAA CTTCCTCTTT GTAGGGTCTA CTGGGGTTGG	1320
TAAGACGGAG CTTGCTAAGC AATTGGCACT CGATATGTTT GGAACCCAGG ATGCGATTAT	1380
CCGTTTAGAT ATGTCTGAAT ACAGTGACCG CACAGCTGTT TCTAAGCTAA TTGGTACAAC	1440
AGCAGGCTAT GTGGGTTATG ATGACAATAG CAATACCTTA ACAGAACGTG TTCGTCGCAA	1500
TCCATACTCT ATCATTCCTCT TGGATGAAAT TGAAAAGGCT GACCCTCAAG TTATTACCTT	1560
TCTCCTCCAA GTTCTAGATG ATGGTCGTTT GACAGATGGT CAAGGAAATA CAGTAACTT	1620
CAAGAACACT GTCATTATTG CGACCTCAAA TGCTGGATT TGGCTATGAAG CCAACTTGAC	1680
AGAAGATGCG GATAAACCAG AATTGATGGA CCGTTTGAAA CCCTTCTTCC GTCCAGAATT	1740
CCTCAACCGC TTTAATGCAG TCATCGAGTT CTCACACTTG ACTAAGGAAG ACCTTTCTAA	1800
GATTGTAGAT TTGATGTTGG CTGAAGTTAA CCAAACCTTG GCTAAGAAAG ACATTGACTT	1860
GGTAGTCAGT CAAGCGGCTA AAGATTATAT CACAGAAGAA GGTACGACG AAGTCATGGG	1920
GGTTCGTCTT CTCCGTCGCG TGGTTGAACA AGAAATTCGT GATAAGGTGA CAGACTTCCA	1980
CTTGGATCAT TTAGATGCTA AACATCTGGA AGCAGATATG GAAGATGGCG TTTTGTTAT	2040
TCGTGAGAAA GTCTAAGACA GAATTTTGAG GATAAAAAAG AAGGAGCCAG CTGAAAAAAA	2100
CTGGTTCTTT TTTAGGTACG ACAGGCATGT CGTATAGTAG AACTGTATTA TTCTAGTTTC	2160
AATATACTAT AGTAGCTCAG AAGTCGGTAC TTAACCTGTC TATATCAAAA CCAGTCCTGG	2220
AAAAACGTGG ACTGGTTTCG TGTTTGGATT ATTACCTTGA ACGACATGCG TTAAGGTTA	2280
GTTGAACCGC CGTATGCCGA ATGGTACGTA CCGTGGTGTG AGAGGGGCTA GAGATTATCC	2340
CCTACTCGAT TTAAATCAC ATGACGTTCA AAGGCATCAT CTGAAATCCC TTGTTCCAAG	2400
ATGAGTTTGG CCCATTCTTT AGCAGAGAAG AGGCTGTGGT CCTTGTAGTT TCCGCAAGAT	2460
TCGATGGTTG TCCCTGGGAC ATCTTCCCAA GTAGTAGTTT CAGCGATTTC CTTGAGCGAA	2520
TCCTTGATAA CAGCTGCGAT TTAGCACTG GTGTGACGTC CCCACATAAT CATGTGGAAG	2580
CCTGTGCGGC AACCAAATGG TGAACAGTCA ATCATGCCGT CAATGCGGGT ACGGATGAGT	2640
TTGGCTAAGA GGTGCTCGAT AGTGTGAAGG CCGGCAGTAG GGATAGAGTC TTCGTTTGGT	2700
TGCACCAAGC GAATATCATA ATTGGAGATG ATGTCTCCTT TTGGTCCTGT TTCTTCCCA	2760
ATCAAGCGAA CATAGGGTGC TTTGACAATG GTGTGGTCAA GTTCAAACT TTCGACAATA	2820
ACTTCTTTTG ACATGGTAAA TCCTTTCAGT TTTCTTCTCT CATTATATCA TAAAGGTTGC	2880
TCCTGAGACA GAGAGAAAAC CTCTCCGAGG CTGGAGAGGT TGAAATCTTT ACTTACGATA	2940
TAAGCGGTCG TATTGGTAGT ATGGGTCAAA GGTACGTTG ATACCCAGTT TACGAAGGAC	3000
ATCTTGTCT TCATCAGTCA AGATGATGGT TGAGTGGGCT TCGCTTCCTT TGAGGTTGCC	3060

679

GAGTTCTTCC ATAGCGCGGG CAGCATCAGG ATTTTCTGTA GCTGTGATAG CAAGTGCAAT	3120
CAGGATTTCA TTTGAATGAA GGCCTGGATT GCGGCTACCG AGATGATCGA TTTTAAGACC	3180
TTGGATTGGC TTAACAACTT CAGGCTCGAT TAGTTTACT TCTTTAGCGA TGTCAGCTGA	3240
TTTTTTGATG GCGTTGATCA AGGCAGCGGC TGTAGGACCA AAGAGTTCTG AGTTCTTACC	3300
AGTGATGATT TCCCCATTG GCAATTCAAA GGCTAGGGCT GGTCCACCAG TTTCTTCTGC	3360
TTTTTGGCGC GCAACGACAG CAACCTTACG GTCTGCAGGT GTGATACCGA GGTCTTCAT	3420
GAGCAACTCA ATTTTCTTGA CGGCAGCTTC GCCAACTTTT TCAGCTTTGA AGTCAAGAAC	3480
TGTTTGATAG TAACGGCGGA TGATTTCTTG TTTAGAAGCT TCGACAGCGG CCTCGTCATC	3540
TGTAATAGCG AAACCAACCA TGTTGACACC CATATCTGTC GGTGAAGCGT ATGGTGATTT	3600
TCCGAGAATA CGTTCCAACA TCGTTTTGAG CACTGGGAAG ATTTGATAT CACGGTTGTA	3660
GTTGACAGTG GTTCTCCAT AGGTTTGAAG ATGGAAGGGG TCAATCATGT TGACATCATC	3720
AAGGTCAGCT GTGGCAGCTT CATAAGCCAA GTTAACTGGA TGATGAAGGG GAAGATTCCA	3780
AACAGGGAAG GTTCAAATT TAGCGTAGCC AGATTTGATG CCATTGATTT GGTCTGGTA	3840
CATATTGGAC ATACACGTTG CCAATTTTCC AGAACCAGGT CCAGGAGCGG TTACGACAAT	3900
CAAGTTGCGA CTGGTTTTGA TGTAGTCGTT TTTGCCCATG CCTTCTGGGG AAATGATGTG	3960
ATCCATATCC GTCGGATATC CTTTGATTGG ATAATGAAGA TAAGAATCAA TTCCGTTTTT	4020
CTCAAGTTGA TTGCGGAAGG CATCTGCAGC GGGTTGGCCA GCGTATTGTG TAATGACAAC	4080
GGAACCAACA AAAATCCCTA ATTCATTGAA TTTATCAATC AAACGAAGAA CTTCTTGGTC	4140
ATAAGAAATG CCTAAGTCGC CACGTGCTTT GGAATGTTC AATGTTGCTAG CATTAAATGGC	4200
AATCACAACC TCAACCTGCT CTTTCAATC TTGCAAGAGC TTGATTTTGT TGTCAGGTTT	4260
ATAACCAGGA AGGACACGAG CAGCGTGGAA ATCTTCTAAC ATTTTACCGC CAAACTCTAA	4320
GTAGAGCTTG CCGTCAAATT GGTTAATGCG CTCGAAATA TGGTCGCGTT GTAAATTCAA	4380
ATATTGTTCA GAACTAAAAG CTTGTTTTTT CATTTTTTTA CCTCTGGACT CTATTATAAT	4440
AAAAAATTGG AAGTTAGGAA ACTACGGAGC TAAAAAGAA ATTAAGAAAG TTAAGCAAAC	4500
GCTTGACAA AATTTTAAAA AGTGCTATCA TAGACTATAG ATTATGAAAA TAATGAGGTA	4560
AACAGATGCA AGAAAAATGG TGGCACAATG CCGTAGTCTA TCAAGTCTAT CCAAAGAGTT	4620
TTATGGATAG TAATGCAGAT GGAGTTGGTG ATTTGCCAGG TATTACCAGT AAGTTGGACT	4680
ATCTAGCTAA GCTAGGAATC ACAGCAATTT GGCTTTCTCC CGTTTATGAC AGCCCTATGG	4740
ATGATAATGG CTATGATATT GCTGATTATC AAGCGATTGC GGCTATTTTT GGAACCATGG	4800

680

AGGACATGGA TCAGCTGATT GCAGAAGCTA AGAAGCGTGA CATTTCGTATC ATCATGGACT	4860
TGGTGGTCAA TCATACCTCA GATGAACATG CTTGGTTTGT CGAAGCCTGT GAAAATACTG	4920
ACAGCCCTGA GCGAGACTAC TATATCTGGC GCGATGAACC CAATGACCTA GATTCTATCT	4980
TTAGTGGGTC TGCTTGGGAA TACGATGAAA AGTCAGGTCA ATACTATCTC CACTTTTCA	5040
GCAAGAAACA GCCGGATCTC AACTGGGAAA ATGAAAAACT TCGCCAGAAA ATTTATGAGA	5100
TGATGAACTT CTGGATTGAT AAAGGTATTG GTGGTTTCCG TATGGATGTT ATTGACATGA	5160
TTGGCAAAAT TCCTGACGAG AAGGTAGTCA ATAATGGTCC TATGCTCCAT CCCTATCTCA	5220
AGGAAATGAA TCAGGCGACC TTTGGAGATA AGGATCTCTT GACAGTAGGG GAGACTTGGG	5280
GAGCAACTCC AGAGATTGCC AAGTTCTACT CTGATCCAAA GGGGCAAGAA TTGTCTATGG	5340
TCTTCCAGTT TGAACATATC GGTCTTCAGT ATCAGGAAGG TCAGCCTAAA TGGCACTATC	5400
AAAAAGAGCT GAATATCGCT AAGTTAAAAG AAATCTTCAA CAAATGGCAG ACAGAGTTAG	5460
GAGTTGAGGA CGGCTGGAAT TCCCTCTTCT GGAACAACCA TGACCTCCCT CGTATTGTCT	5520
CAATCTGGGG AAATGACCAA GAATACCGCG AAAAACTGTC CAAAGCCTTT GCAATCTTAC	5580
TTCATCTCAT GAGAGGAACT CCTTATATCT ACCAAGGTGA GGAGATTGGG ATGACCAACT	5640
ATCCGTTTGA AACACTGGAT CAAGTAGAAG ATATTGAATC TCTCAACTAT GCGCGTGAGG	5700
CTCTTGAAAA AGGTGTTCCG ATTGAAGAAA TCATGGACAG TATCCGTGTT ATTGGACGTG	5760
ACAATGCCCG TACCCCTATG CAATGGGACG AGAGCAAAAA CGCTGGTTTC TCAACAGGTC	5820
AACCTTGGTT GCGGTTAAT CCAAAATTACG AGATGATCAA TGTCCAAGAA GCGCTGGCAA	5880
ATCCAGATTC TATTTTCTAT ACCTATCAGA AACTGGTCCA AATTGCAAG GAGAATAGCT	5940
GGCTAGTTTG AGCTGACTTT GAATTGCTTG ATACGGCTGA TAAGGTCTTT GCTTATATAC	6000
GTAAGGATGG CGACCGTCGC TTCCTAGTTG TGGCTAACTT GTCCAATGAA GAGCAAGACT	6060
TGACAGTAGA AGGAAAAGTC AAATCTGTCT TGATTGAAAA CACTGCGGCT AAAGAAGTAC	6120
TTGAAAAACA GGTCTTGGCT CCATGGGATG CTTTCTGTCT GGAATTACTA TAAATATTTT	6180
TTGCAGAAAA ATTTAAATTT GAAATCGTAT AAAACAAGG GAGGACTGTA TAAAAGACAG	6240
AAATCCTTTG TTTTTATAA CCAAAGTTTA TAACTTTCA TTCTTGAAAT TCAATTAACT	6300
TTACAAATTC CCACTATTAA GGAGAAAGAA GATGAACATA AAGAAGCGTG TCCTTAGTGC	6360
AGGCCTGACT TTTGCATCTG CTTTGCTTTT ACCCAAATCA TTCATACCTC TCTCAACTAG	6420
ATGTAACCTA CAAAACCCCT GACCTCATGA GCCACTTTCT TCCTCCTCAT GAGGTCAGTT	6480
TTACTTTCTG CTGTTCCAGT ATCGTTTTTC CTCGCTAGAT TTCCTCAAAA GGGCAGACTC	6540
CTCCCTTGGT GCGTCACACG ATTTTTCAT CTCGACTGTT CTTTAATGCA TCATTAACGA	6600



681

CGCTTTTCTT CTAGGTGGTT CATAAGGAAC AGGAAGATTC AGGTTGACTT TTCTAATCCT	6660
AGAATAAAGT GCTGAAAACA ATTCGGAATA GGCATAGAGA CTAGACAATT TGAGGAGCTG	6720
CTTGCCTCCT GTTCGAACAC ATTTTCCGG	6749

(2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1842 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

TCTACCCATG GACTTTGAGG CATTCAATTGT TCCATCTTCT AGTGGCGAAT CTTTGTATAC	60
AAACGATTCA ATTCACCTGG ATAGTGAAAC TCTCCCGCAA ACATTTTTCT GGTAACTCA	120
ATCCAGCTGA TATTTCTTTC AGCCAAAATA ATGGACAAGT TCTCCCAAAA TCGTTCAGCC	180
ATATTGCTTC TCCTTTAGTT AGATAAATAA TGTGTTTGGC CCATGTAAAT CAATTGTTTC	240
GTATCTCTTG GCAATAGAGC TCTAGCCTCT TCCAAATTCA GACTTGGATA AACTCGCTTA	300
TTTGAAACCG CAAGAGGAAG TCTGATGGT AGTTCAGGAT TTTTAAAAAT TATCTCAACG	360
AAATCCGTTA ATCTTAGATT GTCACGGTTC TTAATCGTA ATAAATTGGG AGATAAAAAAC	420
TCAAAACAAT CTGAAGAATA GCTCATCATC TCAATTAATT TGTCTTTGT CATTTCAGAA	480
ACTGAATGAC AAGATACCTC TATGCCATAG TTTTGAAGA AATCTAAAAG AAGTTGATTT	540
CTTTGTCTAT TTTTACTTAG ATAGAGATCA ATCATGGGAG ACCTCCCAAA GATTCCGGTTC	600
CATTTGATAT TCTGACACGA TTAAGGAATC TAATAAATTA AGGAATCTAA TAAATTTGCG	660
AAGTTAATCG GTTTCTGTG TTCATCATAA GCTTTTACAG TTACTTGGGT TGTAAGTATT	720
CCCTCTTTTC CCTCGGCTCG ATAGCCTTGT CCATATAAAA CAAAAACGAG ATTTTGATGA	780
TCATCTACAA AGGCATCAAC CCCATTCTTT ATGTCTTGAC TTTCAAGGAA TTCCATAACG	840
TTTTGAAGAT AGGATTCGTA AAATAGTGGG TAGTTATGTT TTTTATGGTA ATCATCTAAA	900
AATGTCACTT CAAACTCACA TGGAGAGTAA TTTTGACTTT GAACAGCCTA AAAGTGCCAT	960
CAAATTTGAA TTGGAATAAA TCAAATAAAT AGCCCCATCC TCATCAATCC AACCTTTGCT	1020
CAAAGACAAC TCCAACCGAT CTTTAAAAAC TGAGTAAACC ACCTTAACCT CCAGTTTCAT	1080
ATTCTTATAC CGTTCACCTC CAAATAAAAG TTTGGGGAGC TTATAATAAC GCTCTGATGT	1140
CTGATATTGA TTAGCGGTAA TACGCTTCAT TATTGTCCCT CCAAGACTAA AATTCCAACA	1200

682

TTTCCAAATT CATCAAATCG GATTAAACCT ACTTGTTCCTA TTTCATCAAC TAACTGAGTT	1260
GCTTTTACCC AAATCATTCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC	1320
TCATGAGCCA CTTTCTTCCT CCTCATGAGG TCAGTTTAC TTTCTGCTGT TCCAGTATCG	1380
TTTTTCCTCG CTAGATTTC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT	1440
TTTCATCTCG ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA	1500
AGGAACAGGA AGATTCAAGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC	1560
GGAATAGGCA TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCTGTTC GAACACATTT	1620
TCCCACCACG TGAAGAAAAA GATGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC	1680
TCCAGCTAGA TGTGTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG	1740
AATTCGTTT TTGATTAAGG TTGAATATC CGTTTTATCG CCAAAAAATC CCTCCTTCA	1800
CTCCTTGATG AAATTCTCG CTTGACCACG TCCACGATAA AG	1842

(2) INFORMATION FOR SEQ ID NO: 86:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19390 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

TCATCTTTAT CTCCTCGAAA TTTTCTAATA TAGCCATTAT AACAGAATTT TGTGAAAAAT	60
CCTATTATAG TAAATCACTA TTTCAGTATA AAAAGAAAAA ACGAATCAGA CGATTGCTC	120
TTCTTAAAT CTGAAATAG CTTTCCAGAA AGGATTAGCC GATTTTGTG AGATTGAGCA	180
CTGCATCGTG ACTCATCAAG ACTTGACCAT ACTCTTGTA GACTGAGCGA CTGATATCAC	240
TATCGTCTGC AAACGCGC ATACGGGCA ACAGCCAAGC TGGATATGGG CTTGGATGAT	300
TTTCAATATC CACTAAAATG GTCAAATAAT AGCGCTCGTT CATTTTGTAG AGTTCAGAAG	360
TTTCCATTTC AAAAGTCACT GTCTTGCAA AAGCTACCAA GTCAGCCAAC TTAGCAAAAG	420
AAAGGATGTA GTAGATGTAA GGTCTTTCT TACTCTCAGC TTCTTGTTCA GCCTGCTCTT	480
GCTCTTCTTC CTGACTTCA ACTTGCTCAA GAGATTGAAT GGCTTCGATA TCATCCTTGG	540
TTTTGTCTGC GATGCTTTT TCCAGGGTTT TGATAAATC ATCTGGAGAC ATTTGAGCCA	600
ATTCTTCCAT ATCTGGCAAA TCCGATAAGT CTTCAAAATC TAGATTTTGG TCAATCTTTG	660
ACTTGGTCAC AAAGACATCT ACCTTATCAG GTTTTGGAGT CACACGGAAG CTCAACATGC	720
CTGTATCCAG AAAGCTATCA GGCATCTCTA GCTCATCCAA GATAGCATAA AAGAACTCTT	780

683

CTGTTTTTTC TTGAGGAACG AGAAAGTCAG CAATCTCCAT TCCACGATCC ATCAAATCCT	840
CTAAAGATAT CGTGATTTTT AAAGTTGTAT CACTAATTG TTTCATTTTC ATTGCTAGTA	900
ACCTCATACT TTCAGTTCTA TCTATTATAC TAGATTTTTA CGATTTTATC AAAAGAAGGC	960
TCCTCTATAC GGATAGATTT TCCCTAGGGT CTTTCTATAG GAGACTCCAA AAGAAAATTT	1020
CTGCAGACAG ATAGAAAAAG CCTTCAAAAT CGGCTAAGAG CCGACTTTGA AGACCTTATA	1080
CATCAGAATA CTTATAATTT AAAGGTTGCT ACACCGAGGA TAGAACGATT TAAGTTTCTG	1140
AGAATTTGAA GACTTTGCTC AAATTTCTTA TAACGAGTCA CTCCGTA CTC	1200
AGGACTGTAT CTCTTTCCAA AAGAGATGAT ACATCCTGTA AATCTACAAA ATGCATTCCT	1260
TTTAAAGCTT CTTGACTCTG TTTCAATTTA TCTAAGATAG CTTTATTGTA GCTAACGATG	1320
GTCAATTCCT GTCCAGTATT TTTGTATGAC AAAACATCTG CTAGGTTAGC AATTGTTGTA	1380
ATCTCTGTTA CAAAATCAAT TTGATACTGA GAAAAATCAC CTACTCTATT GATTGTTGGA	1440
TTAAAGAGAT AACTAACAC ATTTCCCATC ACAACCAAAA TCACACAAAC CACTCCAATA	1500
ACAATAAAC GAAGAATCAG ATTTTTCACA TTTAAGCCAA GCGCTGTTTC ACCATTTGCG	1560
TTCAATTCCT TAGAGTTGAT GGTTCACAGT TTTTCAATTT TCACATTTGC ATAGGCATGT	1620
TTAAATTTCT CAATCAACCC ATCAATTTTT TTCTCTAACA AGTTATTGGC ATCTTTACTT	1680
GATGTCAAAA TTTTCACACC AACCCTGCA TCGTCAATCA TATAGTAGAC GGTCAATTTT	1740
TTCCACCAAT AGTCATTCGT TGAATTTTTC AAGGTTGTTT CTGTCGTGTC TAATTCACCTG	1800
GCAATTTTTT TCAACTCACT GGGTCTACA TCATTGAAAA GATAAGCTCC ATTCAAATTA	1860
CCATCAATCA ATTTCCCATC AAAATCACTA TAACCACCAA TTTGATGATT CAAAATCGTT	1920
TTGTCCGACT CTTTGGAGG AGTGATTTTA TAGATAAGAT AAGTTGAATA ACTTGTGTA	1980
TCTTTGACAG TGTTTTTATT CCTAACTGCT TTAATTGTAA ATGGTACAGC AATGAGAGCA	2040
AATAAAGCGA TGAGAGCTAA AATATTTGCT TTTGCTTTT TATAAAGATT TGCAAACAAA	2100
TCAGCTACTG AATAATGTTT AAACATGATT TTTTCTCCT TTGTTTAGTA GATACTAGTT	2160
TTCTTTGTA AGCATTTTGT CTACAAATAT AATCACAAGA ACAATCCCC AGAATTGCAT	2220
TGTAATAAAA TTGAAGAAAC TTTCTGAAAA GCTGCTTCTT GGCATAAAGA ATAGATTATT	2280
CAAGATGAGT AGGGATAAAG CAAATAGGAT TGTCTTGAG CGATAGGCTA CTTGCAGCAT	2340
GGCTATAAAT AATACGCCGA GTAAGAACT AAGCAGAAAG ACTCCAATCA TACCATAGTC	2400
GGTATACAAC TCCATGATAT AACTACTTCC GATACCATGC CTTTCAAGT ATTCCTTGTT	2460
CAAGACAAGA TAGGATAGAT TGTGGGCATA ACTATTACTA TCAATAGCTA GTTCCACACT	2520

584

ATTGGTTGTA TGTTCAAAGG CTTTTCCTCC GAAAATGGCT CCCAACTCC CCCTTGCAAA	2580
ATAATCAAGA ACAGGACCAA AAGTAAATTT ACGGAAATCT CGGTAAGGGA GGCTACTGTT	2640
AAATAGAAAA CCTCGAGCCA GAACACCAAA ACTAGTCCCT TGTATTATAGA TAAAGTCAAG	2700
TAAGATATCC CAGAAACCTG TATGGGAAAC TTGGACATTA TCCCGTACAT AATTGAGTAC	2760
TCCCATCGCT AACATGAGAA TAGGAGAACC TACAAAAATC GCTAACTTTT CTTTAAACCC	2820
AATCCATTTT CCTTTTTCAG TTTGCTCCCG CATAAAGTAA TAAACAAAAG CAAATMAAAT	2880
ACTTAAATA AAGGGATTTC GTGTCCCAAT TGCCAAATGA ATAGTATTAG CTGCAATAAA	2940
GGAGACAAGC ACTGCTGTGG CCTGCAATTT CTTTGGCTTG GTTGCCAGAT ACATACACAT	3000
TGCATAGACC GTAAAGCTAG ACAAATGTA GGTAATAA GGCAGTTTAC TTTCAAAATT	3060
TGCATAGTAG GCATAGTAGG AAGTCTGCAA ACGATACAAG AGCCGTTCAA ATAACCGAAT	3120
GAAATAGAAA GGATAAGTTA GAAGAAAAAC TCCTAGTGAT ACAAAGCGTA ACCGCTTGAT	3180
ATAAACCTCT TTTAGAGAAT TTCCTATATT TGCTACTTTT ATTTTCTTCC TAGCTATGAA	3240
GTAACGAGCC AGAATGCCTC CTGTGGTCAA GCCCAGAATC GAAATCATGA CAACTATAAA	3300
GGCAAAACGA TAGGCTATTG GATGATAGGT ATCCAAAGCA CCATCCCTAA AATAATCAAT	3360
GGTCGGTCTT GATACCAGAA ATACAAAAAT GGTAAATAG AAAATAAAAT GGATTAAGTA	3420
ATACTTGATA TCATTCCAAC AAGCAATTAA GCTACTAACC AACAAGAACA ATAAAGTAGA	3480
AAGTAAGCTA ACATTATTAT TATTAAACAG ATACACAATT CCACTTACTA GCGTCAAGGC	3540
ATAACTGACT ATGGTCAAAC TAAATAATAA TCGTTTCCCA TCAATCACTT GGTCACCCCC	3600
GTTCTAATGT AATTTTTTAG ATTTTCAAT ATTTTTCAGT AATAAGAATC GATATAAGGA	3660
AATATTTATG AATAGGGCCA AAGCACTAAT TCTTCTCCCC TTACGGAAAA TTGGATTCTT	3720
AGAAATAGCA AAGGCATGGC CTTTTAAAAA ACGATGAATC TGAGAATAGG CTTCAAACCTG	3780
TTTATACTGA TCATCTAGCA ACATCTTATC CAGAATAAAG AAGTGGGCAT AGGCCAATCT	3840
GAAAAAAGCG ACCTCTTTCA AGTCAGGATA GTTTTTCACA ACTTCATTAT AAAACTTTTG	3900
GTAGATATCA ATATAGGCTA AATCCTTCTC TGCATAGGGT TTGGTCGTAA TACTATCCCC	3960
TCTATGGAAA TAGTAATAAT AGGGTTTAGT ATTAACCACA TACTTCTTGG CCAACTTGAT	4020
TAAATCAAAA TGGTAATAGG CATCTTCGTA AATCAACCCC TTAGGAAAGG ATAGGGCAGT	4080
TGCAATCTGT CTCTTGATTA GCTTATTGCA AATCGTCCCA GGTATTTTTT CACCTATGAG	4140
GTATTCCTTT AGAAATGTTT GAGAATCACA GACAAAATAG TCATCCTGAT TGGCTGACTG	4200
TGGGCTTTCA TCATTAGCAT AGACATTCAT GACACCACAG CTCGAAACAT CCGCATCTTC	4260
TTGAACTAAT TGCTCATATA AGCTCTGAAT CATTTCTGGA TGGATATAAT CATCTGAGTC	4320

685

AATAAAAATC AGATAATCCC CGTGAGCCTG CTTTCATCCCA TCATTTCGTG CTTGCGACAA	4380
TCCTTCGTTT TTTTATGAA GCACTGACAC CCTGTCATCT TGTTCAGCGA TTGAATCACA	4440
CAAGCGACCA CTTTCATCTG TTGCACCATC ATCAACAAGA ATAATTCCA GATTTTGATA	4500
GGTCTGCTTC TGAATGGAAG CTATCGATTT TTCTAGGTAC TGCGCCACAT TATAGACTGG	4560
CACAATCACA CTAATTAATG CAGTTTCCAT GCTACTCCTC TAATAGTTTT TCTACTTGTT	4620
CGATTTGTTT TGAATTGTA AATTGTTGAA TGAATTGGCT AGCCTCATCG ACATCAAAGT	4680
TTGAGGCAGA AGTCATGTAA TTAGTAATCG CCTGAGCTGC CTCTTGATTG CTCTCAATGA	4740
TTTGTCCAAA TCGTCCTTCT TGGGATAATT CCTCAGCCCC TCCAACGTCC GTAGAGATAA	4800
AAGGGAGTCC CAGACTCAAG GCCTCCACAT AACTCCAGG AAAACCTTCT TGTTTAGACA	4860
TAGACAAAAG AACTTTCGTC TGAGATAGAT ACTGATAAGG ATTTTTTTGA TAACCAAGGA	4920
AATGTACATA GTCCTCAATC CCATACTCTT TGACTCGTTT TTTTCAGTCC TCTTCCATAT	4980
CACCAGCCCC GATAAAATAG AGATGATAGT TTTTCCCTC TTGGTGTAAT AATCGTATCA	5040
CTTCCACTAC ACGGTCAGAA CCCTTATTTT CCTCAATCCG TCCGATAGTA CAGATACTTT	5100
GAGGAGCAAT CTCGATATCG ATCTTCTCTT GAGATTTTTC TAGAATAGTC TGAAAATCAT	5160
ATCCATTGTA GATTGTCTGT AATTTAGAAG TATAATCTGG ATAAACTTCC TTGATAGAAT	5220
TGCTGGTCTT TTTTGAAATC CCTACAATTG TATTCGCAGC ATCCAAGTGG CTCTATGTG	5280
ATTCTCTTTT AGAGCTATCC TTAAGAAGTT CTCAATACT TCCATGAATC CAAGATATCT	5340
TCTTGACTTC TCTTCTTTA GAGAACAACA GTGGTGGATT CATAATGGTA AAAGAACTT	5400
CAACATCATA ATCATCTTTT ACAAGCAAAC GACGAGTCAG TCTTGGAAAA TAAATCTCA	5460
TTCTCCACAA AAAAGCTCGT AACCATCTGG TTTGGCGATA ATCTTGAAGG GATTTTAAAA	5520
TGCGTACATG CTTTGAACA GATTCATATC CCTTGTCAA GTGCTCCATT TCAAGAATAT	5580
CAATATCATA CTTTCTGGA TCCAGATTG AAACAATGGT TGATAGAATC TTCTCTGCAC	5640
CACCTCCAAG AGAAAAAGAC CACATAAAAA ATAAGATTTT TTTCTTAGCC ACCATATTCT	5700
CCCTTGATT CTGTATAAGA CTTATCCATA TCAGCGATGA CAGCATCATG ATGCGGTACC	5760
TGCTTGCTG CTGGTGGAGG CGTCATATAA TCCCCAAAAG CAGTTCTGAG ATAGACATCA	5820
TAGCCGATTG GAATAGGCAT CTCTGTTCTT TCAAATGGCA AGAAAAGATT GTCTTCAAAA	5880
GATGTGATTG GGTACTTGTT TCTCATGTAG CCAGGACCTG AGCATAATTC TGTAATGCCA	5940
TCACAATCAG CCAAATCATA CTTAGTCATT TCTTCTCAG CTTTTTTCCA GATGCGATAA	6000
CGGAGAGATT TTGGAGTCAA ACCCAGTAAA ATGCGACTTC CCCATTTTAT GAGATCACCA	6060

586

TGCTTTCTG GAATAGTTTG CGCACAAAAG AGTGAATAAA TCAAGGCCCA ACGAACCTGT	6120
TTTTTCCGCT CAGCTGGATT TTTCGGATAA TAATCCAAAG GCAAAACATC CAAGGCCAGA	6180
CCATGTGGCA AATCCAAATC CTGCTGATAA GGCTTGATAC AGGTGGTTTT CTGTGCACGA	6240
ATGGTAATAA AAAGATTACG ATCAACAAAA TCCTTGAC TCTTTGACAA GAAATAACGT	6300
TCATCTGCAT AACGAGGCCA TAATTCTGCT AATTTCTCAT AATCTTTACG AGGCATAAAA	6360
AAGTCTAGGT CGTCGTCCCA AGGAATAAAT CCCTTGTTTC GAAGGGCACC AATAGCGCCT	6420
CCGCCACAGA GATAACAGAG CAAATCATGT TCTTTACAAA AGGCCACAAA ATATTCAGCC	6480
ATCTCCAGAC TACGAGCCTG AATTGCTTTT AAATCAGTCA TATTGTTTAT TATTCTTTCT	6540
ATCGTATCGT TTCATTATAC CACAAACAAG GGGTGAAAAT CTATTGCAGA CTGTAAAAAA	6600
TCAAAGCCTG ACTGCTATCC AAATAGCTAT CAACTTTGA TTTTCTGTC TTATACTCTT	6660
CGAAAATCTC TTCAAACCAC GTCAGCTTCA CCTTGCCGTA GGTATAGGTA ACTGACTTCG	6720
TCAGTCTTAT CTACAACCTC AAACTGTGT TTTTAGCAGC CTGCGGCTAG CTTCTAGTT	6780
TGCACTTTGA TTTTCATTGA GTATTATCTT ATCTTAAGCC CATTTGAGCG AGCTTGGTTT	6840
GATATTTGTT TTGATCAACC AGCAGGCCCA AGCCCCATA AACATCATAG GCATCTACCC	6900
AGTCACCCAG TTCTGGAATC GTCAATTTTT CAATACCATT TTTTGCTCCA TCCAAAACAG	6960
ATAAACCGTT TGTTAGGAGG AAAGTATAGG GTACGTTGGT TGAGGTCATA GCAAAAACCT	7020
TTCAAAGAGC TTCAGAACCA GTGAAAAGTT TAGTGGGATC TTTAATTTGC TCTAAAATTG	7080
CTGTTAAAAC TTGTTGCTGT CTTTTGTAC GGCCGTAATC TGCCCTCATCA TCATCACGGA	7140
AACGAGCATA ATTGAGCAGG GTCGAGCCAT TCATCTGCTG TTTTCCGACT TTAATGGTTT	7200
GGGTTGGAGA CTCAGTCTCG GTAGCGTATA AATCATCTCC GACTGTAGCT TCTGTTAGGG	7260
GACGCCCATT CAATGTTGAA AATTGAGCAT CAATCGTCAC CCCATCAGGG AAAAGCGTGT	7320
CAATCGCTGT GGCAAAGGCC TGGAAATCAA CCAAGGCGTA GTACTTAATG TCCAAGTCAA	7380
AATTATCTTT CAAGACTTGG CGAACCATTT CTGCCCCCTT TTGCCCTCT TGTTCCTCTA	7440
ACTCGTAGGC TACGTTTAAC TTGTTATCTG TCTGTTTCT ACCATTAATC ACTTGACTAT	7500
AACCATCTAT ATAGACCAAA TTATCACGCA TGAACTGAC TAGCTTCATT TTCTTATCTG	7560
AGCCCCGAC ATTTAATACC ATAATAGAGT CAGTTCGTGT CTCAACACTG TTCTGGCCGA	7620
TTCCACCATC AGTACCCATG ATTAATAATAT TAACTCCATC TCTAGTGTCC TGACCATTAA	7680
AGACTTCTAC TTGAGCTGCC CGGGCATCAG CAGTTTCTT TGCGCTAGCA TCTTGGAAC	7740
CACGCAAAAA CATGAATACC ATGGCCAAAG CCACACAGAC CAAAAGTGAA AAAATCACCA	7800
TAAAAATTCG TTTAAGACGG AGCTTCCGTC TTTCTTTTT TGGAGGGAAA GAGAGTGCTT	7860

687

GTGATTTGGA TTGTGAGCGA CTCGGTTTCG CATAGCTTGG TAAGTCAACC TGCTCTTCTC	7920
TTTCTTGTTTC CAAGCTAGAG CTACTATTTC CCCTAGCAAG AGTTAGCTTT TCTTGCAAAT	7980
AGGCAAACTC ATTTTTTTCT CTCTCATTGA GATAGTGAAT ATTTTTTAGC AAATAATCAT	8040
AACGCAACTG CTCATGATGA CTTAAGGGAT TTTCTTTACT CATCTTCTCT CCTTTCCATG	8100
GTCTGATATT GGATAAATAG GATAGGCACC CAGAATTTTA TACTGGATTG CAATCGCTTC	8160
TAATTCTTTT TGGGCAAAGT GGACCAAGTC CTTATCGGTA TAATCCACAT CGATAATGAA	8220
AAAGTATTCA CCCAGTGCTG TCTTGAGTGG ACAACTTTCA ATTTTGTCA AGTCAATTCC	8280
TCGCCAAGCA AAGGTCGACA GGGCCTTATA AAGTGCACCT GGAAGGTTGT CAGGTAATGT	8340
CAAGGCCAAA CTCATCTTTT CAGTTTGTGC TTGCAAGGGA ATACTAGGCT TTTAGCTCC	8400
TAGAACCCAG AAACGTGTGA AATTGGCTTC CATTTCTGTA ATATCCTCGG CAATCAGTTC	8460
CAATCCATAT TCTTCAGCAG AACTTCTAGG TGCAACTGCT GCAAAGGGCT GGTCTGGATG	8520
TTCCGAAATA AAACGGGCCG CATAAGCTGT ACTAGCTGTT ACCTCGATTT GAGCCTCTGG	8580
ATATTGTTCA TCGATGAATT TCTTTCCTTG AGCCAAGGCC TGTGGATGTG AAAAAATCTT	8640
TTCAATCTTA GTATGGCCTG GAACCACCAT CAACTGCTGA TGAATAGGCT GAACGATTTT	8700
TGCTACTGCT TGGATGTGAG CCTGATGAAA AAGATAGTCC AAGGTTTCAT GAACACTACC	8760
CTCAATAGAA TTTTCAACTG GCACCACAGA ATAGTTCACT AATCCTTGCT CATAAGCCTT	8820
GATGACATCT GTAATGTTGG CAAAAGCCTG CAATTCCTCA TGAGGAAAAG CTGTCTGCAC	8880
AACGTGGTGT GAAAATGATC CCTTGGGACC TAGATAAGCA ATTTTCATCT TAGTTCCTCT	8940
ATAATTTCTT CTGGGCTTAG CTTGGTCACA TCCAAAACCC GACTAGCCAC TTCCTCATAC	9000
CAAGCCTGTC TTTCTTGGAA AATAGCTACT AGTTCTTCTT TGCTATTATT TAGAAAAAGC	9060
GGTCGCTGAT TGTCTTATC AGCTGCGATA CGTTGGTAGA GGGTTTCAAA ATCTGCTCTC	9120
AGGTAGATGT TATCTGTATT AGTCTTGAGT AAGTCACGAT TTCTCTGAGA AATAACCACT	9180
CCTCCTCCAG TTGACACGAC TTGGTCTGTT TGTAGTAAAT CAGCTAGGAC TTCTGATTCT	9240
ACCTGACGAA AGGCTGTTTC TCCCTTTTCA GCGAAAAAAT TCGCAATGGA CATACTAGG	9300
CGATTCTCAA TCAGAGCATC CATATCAAGG TAATTAGGGT CCAAGCCTCT TGCAATAGTC	9360
GATTTTCCAG CCCCCATAAA CCCTAATAAC ACCTTAGCCA TGAATCAAGC TCTCCAAATC	9420
ATCAAAGAAA CTAGGATAGC TGGTATTGAT GGCTTCTGCA CGGTCAAGCT CCACCTCTCC	9480
ATCTGCAACC AAGAGGGCTG CGATAGCTGT CATCATGCCG ATACGGTGGT CACCAAACGT	9540
ATTGACTCTA GCACCGTGAA GAGCTGATTT TCCTTTGATA ATCATCCCAT CTGCCGTAGG	9600

688

AGTAATATCT GCTCCCATAC TATTTAAGGC GTCTGCCACA ACCTGAATAC GGTCTGTTTC	9660
CTTGACCTTG AGCTCCTCAG CATCCTTGAT AACTGTTACA CCTTGGGCTT GGGTCGCAAG	9720
CAGGGCAATA ATGGGCAATT CATCAATCAA TCGTGGAATC AAAGCGCCAC CAATCTCTGT	9780
TCCTTTCAAG TCAGAAGACT CAACAATCAA GGTAGCAGAT TTAGCGACTG GATCGATTTC	9840
AGTTATTTCC AATTTTCCAC CCATGGCAGC AATGACATCA ATAATACCGG TCGGAGTTTC	9900
GTTGATCCCC ACATTCTGCA GCACTAGACG AGAATTTGGA GCAATCAAAC CTGCGACTAA	9960
CCAAAAGGCT GCACTGGAAA TATCTCCTGG TACGACCACC TTCTGTCTCG TCAATTTTTC	10020
TGGCCCTGG ACTGTGATTT TCTTACCATC CACACTTAAA TGACCACCAA ATTGTTTCAA	10080
CATATCTTCA GTATGATTAC GGGTGTACTC TTTTTCGATA ATAACGACT CCCCCTTAGC	10140
TTGTAAGGCT GCAAACATCA AGGCTGACTT GACTTGGGCA GAGGCAATTG GCAACTCATA	10200
ATGAATAGGT CTTAGGTTTT TCGTCCCTTT TAAGCGAAGG GGAGGCAAGT CTCGTTCACT	10260
TTGCCCTGAA ATGCTGACGC CCATTTTTTT CAGTGGAAGG GTCACACGGT CCATAGGACG	10320
TTTGAAAGA CTATCATCTC CAAACATCTC TACTTCGAAA TCTGCACCAG CAAGGACACC	10380
TGAAATCAGG CGAATCGAGG TGCCAGAATT TCCCATATTA AGGGCATTTT GTGGCGCTTT	10440
TAAGCCAGCC ATGCCTACAC CTTGAATGGT AATAACCCCA TCTTTATCCT CAATTTCAAC	10500
ACCAAGGTCA CGAAAAACCT GCATGGTCGA AAGAAGCTCT TCACCTCGCA GAATATCATA	10560
AACCTTGGTC TCACCCTCAG CCAAACCTCC AAAGATAATG GAACGGTGGC TGATAGACTT	10620
GTCACCTGGG ACGCGGATAC TACCATGTAA ATGGCGAATG TTTGTTTTTA GTTTCATACT	10680
GGACCTCATA CTTGCAATAC TTTTACCTAT TTTATCATAA AAAGCCAGAA ATTCTTAAA	10740
AATTCCTGAC TTTAGGATCG TTCTTTTCTT ATTTGAGCAA TTCTGAAACT GGTTCAAAA	10800
CAATTTTTTC AATATCAGAA AGGTAAATGG CCAATTGTTG TTGCTTGCTA AAGAATTCTG	10860
ACAAGAGGCT ATTTCTTGA ATCTGTTTAC CAAAGCCTTC CATCTTAGCT TGGAAGGACG	10920
CATCTGGCAT TTGACCTGTC TGTGCTAGTT TTTGAATTC CTCTTGAAAG GCAAGATAAT	10980
CTGTAAAGAT TTTGCTTGGC TCAGCATCTG CTGCAATCGC ATCTTTAGCT GCTTTAACAG	11040
CCTTGATTC TGGTAATCCG CGTAGACCGC GACTGAGTTC GTTTCACCTA TCGTAAATAT	11100
TTGACATGTT CTTCTCCTTA TTTGATGACG ACTGTATAGT CAGTATTTTC TGTTATGAGA	11160
TGCTCAGCTC TTTCCAAGTC TTGAGCATTT TTAATGAAA TTTGTAGGAT TCCGTGAATA	11220
TCCTCAGGAT TTTCTCGTT GATGTGGATA TTAACCAAGG AAGTTCCACG TAGCAGTTCC	11280
AAAATCCGCA GGATGACATC TTCTTCATCA GGAACGTCAA CATAGAGGTC GTAAGAGCTA	11340
TCCACACCAC CACGCTTATG GATTTCCATG GTCTGGCGTT GTTCACGCGC TTGGTTAAAA	11400



689

AAGTTCCAAA TTTGCTCTTC ATCTCCCTTA CTAATGGCCT GACCAATCGC TTCCAAACGT	11460
TCCTTGAAAT CCTCAATTCT ATCCAGAATG ATCTCGCTAT TGGACAAGAG AATGGAGGTC	11520
CACATTCTCG GCTCGCTTTC CGCAATTTCG GTCATATCTC GAAAACCACC TGCCGCAAAG	11580
CGCCTTGCCA TCTCATGCTC TTGAGCATAG ACCGCAGTCT GCTCCATGAG ACTAGAAGCC	11640
AAAAATAGAG GAAAATGGCT AATCTGAGAA GTGACACGAT CATGCTCCTT GGCATCAATC	11700
TCGATAAAAC GAGCATGAAG ACCTGAAAGC AGATCCTTCA TTTCTTAAG CGTGTCTGA	11760
CTTGTCTAGGC TTGAAGGTGT AAAGATATAA TAGGCATTTT CAAAAAGATT GACATCTGCC	11820
GAAGCAGCCC CTGTCTTGTG ACTACCAGCC ATGGGATGGG CCCCACAAA GCGAACAGAC	11880
TTGCCAGCCA AATACTGCTC CGCCGCATCC ACAATGGTTG ACTTGGTCTG ACCAGCATCT	11940
GAAATAATAA CGCCTTCTCG CAAATCCAAA TTGGCCAACT CCTTAATGAA AGCAATAGTT	12000
TGTTTGATTG GCAAGCTGAG GATAATGACA TCTGCCAAAG GAGCAAACT AGCAAAATCA	12060
TCCGTTGCAC GGTCAATCAT ACCTTCTTTC AAGGCGATAT CTCTCGAAGC TTGACTACGA	12120
TTATAACCTA AAATTTTATA ATCTGGATGA TCGCGTTTGA TACCAAGTGC CATAGAGGCT	12180
CCAATCAACC CAAGACCTGC GATATAGATT GTTTTTGCCA TAGGAACTCC TTAATAGTTC	12240
TTGTATAGT CTCGGTGTTC GGCTACCGCT TCTTTTAGTT CCTCAAGATT ATCTGATGAG	12300
AATTTTTCGA GGATTTCTTG CGCCAGAACC GTTGCTACAA CTGCTTCCAT GACCATTCTT	12360
GCAGCTGGAA GAGCAGTCGG ATCACTTCTC TCCACGGTTG CTTTGTAAGG TTCGTGGGTT	12420
TCGATATCCA CACTCATAAG AGGTTTATAA AGAGTAGGAA TGGGTTTCAT GACCCACGA	12480
ACAACGATGG GTTGCCCAT AGTCATACCA CCTTCAAAAC CACCTAGATT ATTGGTACGG	12540
CGAGTATAAC CGTCTTCTTT AGACCAGAGA ATTTTATCCA TAACTTGGCT GCCTTTACGA	12600
TAACCAGCCT CAAAGCCAAG ACCAAATTCC ACCCTTTTAA AGGCATTGAT AGAGACAACA	12660
GCTTGAGCCA ATCTTGATC CAATTTTCTA TCCCATTTGA CATAGGAACC AAGACCAACT	12720
GGAACGCCTC CGACGACTGT CTCCACAACC CCACCGATGG TATCACCATC ACGTTTGATT	12780
TGGTCAATAT AGTCCTTGAT TTCCTGTTCT CGTTCTTGGT TGACAATAGA AACTTCAGAC	12840
TGGGCAGCTC TTTGCTTAAT TTCAGCGACT GTCAGATTTT CAGGAACATC GATTTCCTTG	12900
CCACCAAAGA CCACGACATG GTTGCCAATC TCCATATCCA GCTCAGCCAA GAGGCGTTTG	12960
GCTACTGCAC CAACTGCCAC CCGCATGGTG GTTTCACGAG CTGATGAACG CTCCAAAGAA	13020
TTTCGCAAAT CATCAAAACG GTACTTAATC CCCCCAACCA AATCGGCATG ACCTGGGCGA	13080
GGATGAGTAA TTTTCCGCTT GCTTTTAAGG CGGTCTTCAA TGTCTCCGC AGACATGATC	13140

690

TCCAGCCATT	TCTGGTGGTC	CTTATTGATG	ACATCCATAG	TAATAGGCGC	CCCTGTCGTC	13200
TTCCCGTGGC	GAACGCCCCG	AGTAAAGACA	ACCTGGTCAT	TCTCAATCTT	CATACGACCA	13260
CCACGACCGT	AGCCACCCTG	ACGGCGTCTA	AGGTCCCTCAT	TGATATCCTC	AGCTGTCAAT	13320
GGAAGTCCAG	CTGGAATTCC	CTCAATAATA	GCTGTTAGAC	GGGGGCCGTG	TGATTCTCCT	13380
GCAGTTAAAT	ATCTCATACA	CTCTCCTTAT	TTTACCAAGT	AGTCTTTCAT	CTCTTCCAGA	13440
GAAACTGGGT	GAATGGTCGC	TGAACCAAGC	TCTGGCACCA	AGACCAATTT	CAAGGTGTTA	13500
CCACGCGCTT	TCTTGTCATG	AGTAAGAGCC	TGATAAAGCT	TGCCAACTTC	CCAATTTTCA	13560
TAGTCAACAG	GCAAACCGAA	TTTCTGACAC	ATCTCTGTGA	TAGATTGGGT	AATGCCAGCT	13620
GGCATGAGGC	CTTTTTCCTC	AGCAACCTTG	GAAATCTGTA	CCATTCCCAT	GGCAACAGCC	13680
TCTCCATGCA	TGACCTTGCC	ATAACCGGCA	GTCGCTTCGA	TGGCATGGCC	AATAGTGTGG	13740
CCAAAATTGA	GGTAAAGACG	AATACCATTG	TCCAACTCAT	CTTCAACCAC	CATCTTGCGC	13800
TTCACCTGAC	AAGAATGTTT	AATCAAGGTC	TCTGCATGTT	CCAAAATACT	CTCAACAGAA	13860
CCATTGAGTC	CCGTCAAGAG	AGCCACAGT	TCTGGATCCT	CAATCAAGCC	ATACTTGATA	13920
ACTTCACCCA	TCCCTTCAAT	CAACTCTCTT	TTTCCGAGGG	TTTCAAGAAC	AAGTGGATCA	13980
ATCAGAACCC	CATCTGGTTG	GGCAAAGGTC	CCCACCATAT	TTTTAGCAAA	TGGTGTATTA	14040
ACGCCCTGCT	TTCCACCGAT	AGAAGAATCA	ACCTGAGCTG	TCAAACCTAGT	CGGAATCTGA	14100
ACAAAGTGAA	TACCCCGCAT	ATAGGTAGAG	GCTACAAATC	CAGCCAGGTC	CCCAACAACG	14160
CCACCACCAA	GAGCAACGAT	TCCATCGCTA	CGAGTCAGAC	CTTGCTTGAC	TAGAAATTCA	14220
TAGACTTTCT	GAACAGTAGT	TAAATTCCTT	CTTCTTTCAC	CTTCTAAGAA	ATCAAAAACA	14280
GCTACCTGAA	AACCAGCATC	TTCTAGGCTG	AGCTTGACCT	TCTCTGCATA	GAGAGAGGCT	14340
ACATGGTTAT	CTGTCACAAT	GACTACCTTT	TGCGGTTGCC	AGAGTTCTCG	CAACCACTGA	14400
CCAGCCTGGG	CCATACAACC	TTTTTCAATC	TGAATATCAT	AAGGATGGTG	AGGAATATCG	14460
ATTCTGATTT	TCATAGGAGA	GTCTCCCTTT	CTTTATTGGT	ATTTTCTGT	TAAAGACTGC	14520
CAAATCTCTT	CTGTCGGCAT	TTCCTTGCCT	GTCCACAGTT	GAAAAGCTTC	TGCAGCTTGA	14580
TAGAGTAACA	TTCCAGACC	ATTGACTGCT	GGATTGCCCT	GACTTCTAGC	CCATTTCAAA	14640
AACGGTGTTT	CAAAGGGTTG	GTATATGATA	TCTGCAACTA	AAAGAGTTTC	TGGTAAGACT	14700
ATGTTTTTCAG	GAACAGGAGA	GGATTGGCCA	TCCATGCCCA	CACTGGTGGC	ATTAACCTAGC	14760
AAATCCGACT	CGGCAATCCT	TGCTTGCACT	TCAGAAACAT	ATTCTAAAGC	ACACAAATCC	14820
ACTTTAAAC	CTGTCTGCTC	CTGTAACCTG	TCTAGGTAAG	GTCTTGTTTT	TTCCATAGAA	14880
ACGGAACGAA	CAAAGACCGA	AATCTGACTG	ACGCCATCCA	AAATAGCCTG	TGCCAAGATT	14940

691

GATTTAGCCG CACCACCTGC ACCCAGCAGG GTCATCTTTT TACCTGAAAT TGTA AAAAGAA	15000
GGCAAGCACT TAAAAAATCC CTTGCCATCT GTATTATATC CAATTAAATT GCCATTCTCA	15060
TTGACAACCG TATTAACCGC ACCAATCAAG CGCGCTTCAT CGCTCAGCTT ATCCAAATAA	15120
GGAATCACCT GCTCCTTATA GGGCATGGAC AGATTGATGC CAAACATCTG GTAGCGACGA	15180
ATATTGGCCA CTGTTTCTAC CAAGTCACTC GCTTCAATCT CCCAAGCCAC ATAAGCACCG	15240
TTGGTAGCTG TCGCCTCAA GGCTCTATTG TGGATGAAGG GAGAAATAGA ATGCTTAATA	15300
GGATTGGCAA CAACTGCAGC TAAACGTGTA TAGCCATCAA GCTTCATCCA AAATCTCCCT	15360
GATTTTTTTC ATGCTAGCTA GAGAAATCTG CCCAGGGGCA CTAACCTCAT CCAGACTGGC	15420
AAAAGACCAA CTCGAACCAG TCACATCCGC AGTGATACGA GAGACCTTGC CCACCTTACC	15480
CATAGAAATG GTCACATATT CCTGTTCAGG ATTGAGGGTT TTAAAGCCTC GTGTATAGTT	15540
CATCAAGTCT AAGACATCCT GCTCCGTGTG AGCCATCACC GCAACCTTAA CAAGTTTGTG	15600
ATTTAGGATC GTCAACTCTG ACAAGATTTC CATCATGTTT TCAGGTGTTT CTGGAAATT	15660
ATGGTAACTC AAAACAAGAT TTGGGAAGTC CAGCATTTCC TCAAAAACAT CCTTGTAGCT	15720
ATAGTACTCA AAATCAATAT AGTCTGGTTG ATAGAGTTGC GCAACTTCCT TGATTAGATG	15780
GATATACTCT TCTGGAGAAA GGTCGATTTC TCCACCTTCG GAGCGAGTTC GTAGCGTGAA	15840
AACCAACTCA CGGCCTGCGA ATTTTTCAAA AATGGCTGGA GCTACCTGCA AAATCGCTTC	15900
TTTAGGCAGA TAGTCGGCAC GCCATTCAAT GATGTCGGCA TCCAGGTACC TCGTGGCATC	15960
CAGAGCCTGA GCCTCTCTA AACTTCTTGG CATTACTGAA ACGATTAAAT TCATTTACTA	16020
ACCTTCATAC TAATCACCTT GAGGTAATTA CTACTTTCAT CTTTTTTATT ATAGGCAAAA	16080
TCTGCTGGAA GACCATATTT GTTTAAAATC TGGTAACTTC TTCTTGCAA ACCTTTATCA	16140
ATTTGTCTG TAAATTCTG ACGGGAAACA TTGGCAGCAT TGGTACTGGC AATGATAATC	16200
CCTCCCGGAT TTAAATCTC AAGACTCTGG GAAATCAACT TGTGATAATC CTTGGCCACA	16260
GAGAAAGTTT GTTTTTTATT CCGAGCAAAG CTAGGCGGAT CTAGGACAAT CACATCGTAG	16320
GTCAAGTCTT TCGTTTTGGC ATATTTGAAA TACTCAAAGA CATCCATGAC TATAAACGA	16380
TGCTCGTCTG TGCTGAGCCC ATTTGCCTGA AAATGCGCTT GAGACAATTC TCGTGAACGT	16440
TTGGCTAGAT CAACAGAAGT TGTATGGCTA GCTCCTCCCA TGGCCGCAGC TACTGAAAAA	16500
GGCGCTGTGT AGGAAAACAT ATTGAGTAAG GATTTACCCA TAGCCAAGCC GTCAACTAAA	16560
CTACCGCGAA CCTCATGCTG GTCTAGGAAA ATTCCTGTCA TCAAGCCATC ATTCATAAAG	16620
ACTTGATACA GGACACCATT TTCTAAAACA TTGAAAAAGT CAGGTGCTTC TTGACCATAA	16680

692

ACATGGGCAG ATTCATAGTC CAAACCCCTA AAGCGGATTT TCTCATAAGC TCCTAAAACC	16740
TCAGGGAAAA CCTGTCTAAA GGCTTCTGAT ATAGTCTGAC GAATCTGATA AACATAAGAG	16800
TTATACCAAG AAAAGACGGC GTAGTCGCCA TAAAGGTCCA CTGTCAGACC CCCAAAGCCA	16860
TCTCCCTCTT GATTAAAGAG ACGAAAGGCA GTTGTCAAAT CATCTTGATA GTAGGCGTTT	16920
CTCTTTTCTT TGGCTTTTCT AAACAACGTT TCAAAGAAAG CTTGATTGAA GGCCACCTTG	16980
TCTTTGCTGA TAAACCAGCC CAAGCCCTTG TTTTGCTGAG AAAGGTAGGC AGTCCCAAGA	17040
AAGTTTCCTT CCTGACCCTG CACCTCTACT TCCTGATCCT TAAGATTGAC ATTCTCAAGA	17100
TCACTGGCTT CTAGTAAAC TAGCCCCCTA GCAAGCTTCT TTTCAACCCT TTTGCTGACT	17160
CTTATTCTAT TCATAACTAC CATTATATCA AACTTTTAGA CAATTCTCAA AAAAGAACT	17220
ACCCTTGCTT TTTTACTCTT CTTTAAAAA ATGGTATACT AGACTTCCTG CAAACTAGG	17280
AAGTAAATGT GTAAGAATCA CAGTAAAAA TGCTCTTCCG TCTTGAGGA GCATTCTTT	17340
TTATCAACGA AAATCAAATA GCAAACTATG AAAGTAGCCT CAGGTAACT GTGAGATTAT	17400
AGGTAGAGAG GTTGTATCAG CAATATGTGT CTGTCAAAT TAGTGACAAA GGTAGTAGAA	17460
GAAAGATAAA GAAATAAATC AGCTTCAGTA GGTATCTGGA AAATTGATT TTATAGAGAA	17520
GCCTTTTGTT ACAAACCTCA TATACTATCA ATAAATAATA TTATAGAAGC AACAATAATT	17580
ATAATTTTAC CTATCTGCAT CATTCTATTT CGAACTCTAA ATATATGTTT TATCAAAAAT	17640
ACTTGGAACA CACACATTAT AGGAATTAAC GTTTTGTAAA TTGAAAAATA TCCAAATAAA	17700
TAAACTATAA ACAACAAAAA TAGAACTATG TTATATTTCT TATTCAAAAC ATTCTCCCT	17760
ATATATTTTT GATTACCAAT CTTAATCATT TACAACCTACA TTCTAACAAA CTATAAAGC	17820
GTCTGTCGAA TTGAATTTAT CAAGCAAGCG ACCCAACAGT TCATCTTTTT TCTATTTCTG	17880
CCAATATGCG TGACAGGTAA TAATGATAGC CAAAAATAGC AAGAGCAAGC AAGACGATAA	17940
GAGCTCCTAC TCCAAGCTG ATGGCAAGGA TAGGGGAGAG AGACTGAACC AAGAATATGC	18000
TCCAATTAC AAGGGCCATC AGGATTGCAC TATAAATAAA CAATAAACT ATGGCGAC TA	18060
TGCCATTTGA ACGATTACAC AGGTCCGTAA TGCTACTCCA ATTGGTTGAC AGATTTTTAA	18120
CGTCCTTAAA GTAATGGTGG CAAGAAAGGA TGACACTGGC AATGATCCAG ACTACAAGAA	18180
GGTAAATCAT CGAAATGATG GGCAAGCCTA GATATAGAGA AAGACCAAGC AAAGTCAGAA	18240
CTGGTAAAAA GGACTGGACA GCATATATAA TCCAAAATTT CACTTTCACA TAACGAGCAA	18300
AGTCAAAGGG TAAACTCTTA AGAAAATCAA CATTTTCCCT CTCCAAGGAC AAGGCAATTG	18360
AATGCAGGCT GGTGATATTG TTATTGACAA CTGCTATAAA GAGAGCTATA AAAACAAGG	18420
GTAACCAGTA TGGAGGATGA ATGTCTGGAA CTATCTGAGA ATCTCGGATT TTGGAAATCA	18480

693

GACCGATCAT CATGAGATAA GGAAGGAAAG CACTTGTAAG AAGCACTGTA ATCACGCCAG	18540
TCCCCTGTCC CAAGAGGGTG AGGTGGTAGC GTAAAACCAT GCGAAAAAAT CCCTTTTGTAG	18600
TGTTTGAAT TCTCTCCTTG CTGCGACGTT CTTTTTTGAC CTTCTCCTCA CTATTAAGCA	18660
GGATCACGTC ATAAAAACGA GGAAGGACCT TCTTTTTGGT CAGATAAAGC AGGAAGAGAG	18720
TTAGTCCTAT CCAAGCGAGC AGACCCACTA AGGCTTCTGT CGAAAAAGGC TCCACTGCTA	18780
TTTTGTAAAA GATATGAAGA GGATAAAGGA GAAATGGAAT GTCTCTAACT TTGTCAACAA	18840
TACTTCCAAA AGTCGACTGA AGAAAGAAGA TAAATATTAA AGGTATGAGA ACTCCTATCC	18900
CAATCATCAC ATTCGAAAAA ATAGACTGAT ACTTTCTGAA GACCCTAGTT TGAGCCAAGA	18960
AATGCACTGC CACTACCATC ACTAGAGCCA CAGAGACAAA TAATAAGGTC AAGGACAGTA	19020
GCATCAAAGG CAAACCCAGC CATAGAGAAG GAGCTAGCCT AATGTAGAGG ACCAGAAAAT	19080
AAGCTAGGAT TGGTACAATT CCAGTTAGAG CTGGCAAAAG GACAGACAGT CCTTTAGCAA	19140
TTATAATCTC TGATTCTTTA AAGGCATAGG GCCTATACGA TACCAAATCC TTACTCTCAT	19200
AAAAGACATT GTAAAAGGCC GTTAAAGAAG TTGAAAAGGC AATCACTAGT AAAATAGCAA	19260
TCATCGAGCT AAAATAAATA GGTATTTCTT CAAAAGGAAA ATGAATGGCT ATATTACTAA	19320
AACAGATGAT CATCAAGAGA CTGGAAAAAA TGTAAGAACT TAAGACTCTA GCGGAAACAT	19380
TTACTTTTTT	19390

## (2) INFORMATION FOR SEQ ID NO: 87:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18436 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

CCGAGCGTCG TTACAGACTT TATCAAGATT GGACGCAAGA AGAAATTCAG CATATAAAGG	60
AAAAATATGGC ACAATCTCCA TGGCATACTC ATTACCATGT TGAGCCAAAA ACAGGACTTC	120
TCAACGACCC AAATGGCTTT TCTTACTTTG ATGGCAAGTG GATCCTCTTT TACCAGAATT	180
TTCTTTTGG TGCAGCCAC GGTTTAAAT CTTGGGCACA GCTAGAAAGT GATGATTGTA	240
TTCACTTTAA AGAACTGGA ATCAAAGTTT TACCAGATAC TCCATTAGAT AGCCACGGTG	300
CCTACTCTGG TTCTGCCATG CAATTGGCG ATAATTATT CCTATTTTAT ACAGGAAATG	360
TTCCGGATAA AAATGGATC CGTCACCCAT ACCAGATCGG TGCTTTGATG GACAAGGAGG	420

694

GTAAGATTAC	AAAGATTGAC	AAGATCTTGA	TTGACCAGCC	AGCAGACTCT	ACTGACCACT	480
TCCGCGATCC	ACAAATTTTT	AACTTTCAGG	GTCAATATTA	TGCCATTGTC	GGCGGACAAG	540
ACTTGGAGAA	AAAAGGTTTC	GTTCGTCTCT	ACAAGGCTGT	CAATAACGAC	TACACAAACT	600
GGCAAGCAGT	TGGCGACCTT	GACTTTGCTA	ACGACCGTAC	TGCCTACATG	ATGGAATGTC	660
CTAATTGGT	CTTTGTAGAG	GAACAACCTG	TCCTTCTCTA	CTGTCCACAA	GGATTGGATA	720
AGAAAGTTCT	AGACTACGAT	AATATCTTTC	CAAATATGTA	TAAGATCGGG	GCTTCCTTTG	780
ACCCTAAAA	TGCCAAAATG	GTAGATGTGT	CTCAACTTCA	AAACATGGAT	TACGGTTTCG	840
AAGCCTATGC	AACTCAAGCC	TTCAACGCTC	CTGATGGGCG	TGCTCTAGCA	GTTAGCTGGC	900
TTGGTTTGCC	AGATGTTTCT	TACCCATCTG	ACCGTTTTGA	CCACCAAGGA	ACCTTCTCTT	960
TGGTCAAGGA	ACTCACTATC	AAAGACGACA	AGCTCTACCA	GTATCCAGTC	GCTGCTATTA	1020
AGGACCTTCG	TGCTTCTGAA	GAAGCCTTCT	CAAACCGTTC	CCAAACCAAG	AACACTTACG	1080
AACTTGAAGT	CAACTTGGAA	GCTAATAGCC	AGAGCGAGAT	TGTCTTACTT	GCTGATAAAG	1140
AAGGTAAGGG	ACTTTCAATC	AACTTTGACC	TTGTAAACGG	TCAAGTAACA	GTGGATCGTA	1200
GCCAGGCTGG	AGAACAGTAT	GCCCCAAGAAT	TTGGGACAAC	TCGTTCTTGC	CCTATCGAGA	1260
ATCAGGCTAC	TACTGCTACA	ATCTTCATCG	ATAACTCTGT	CTTTGAAATT	TTCATCAATA	1320
AAGGAGAAAA	AGTATTTTCT	GGTCGTGTCT	TCCACATGC	GGACCAAAAT	GGTATCCTGA	1380
TTAAATCTGG	AAACCAACT	GGAACCTACT	ATGAATTAGA	TTATGGTCGC	AAAACCTAAT	1440
GATGTGCGCA	AACTTGCAGG	CGTCAGTCTT	ACTACCGTTT	CTCGGGTTAT	CAATAAAAAA	1500
GGGTATCTAT	CTGAGAAAAC	CATCCAAAAA	GTCAATGAAG	CCATGCGAGA	ATTGGGCTAT	1560
AAACCAACA	ACCTGGCTCG	TAGTCTGCAA	GGAAAATCAG	CTAAGTTAAT	CGGCTTGATT	1620
TTCCCAATA	TTTCAATGT	TTTCTATGCA	GAATTGATTG	ATAAATTGGA	ACACCAACTC	1680
TTCAAAAATG	GTTACAAGAC	CATCATCTGC	AACAGTGAAC	ATGATTCTGA	GAAGGAACGC	1740
GAATACATCG	AAATGTTGGA	AGCCAATCAG	GTGGACGGCA	TCATTTCTGG	TAGTCACAAC	1800
CTAGGAATCG	AAGACTACAA	TCGTGTGACA	GCGCCGATTA	TTTCCTTTGA	CCGAAACCTA	1860
TGCCCAGACA	TCCCTGTCGT	CTCCTCTGAC	AACTATGCTG	GTGGGGTTCT	TGCTGCCCAA	1920
ACCTTGGTCA	AGACAGGTGC	CCAGTCTATC	ATCATGATTA	CAGGGAATGA	CAATTCTAAT	1980
TGCCCACCG	GACTGCGCCA	CGCTGGTTTT	GCATCCGTAC	TCCCCAAAAG	TCCTATTATC	2040
AATGTTTCCA	GTGACTTTTC	TCCCCTCAGA	AAAGAAATGG	AAATCAAGAA	TATCTTGACC	2100
CGGGAAAAAC	CAGATGCCAT	TTTTGCTTCG	GATGATTGTA	CAGCTATTCT	GGTCATTAAA	2160
ATCGCTCAAG	AATGGGCAT	TTCTGTCCCA	AAAGAGCTCA	AGGTCATCGG	CTATGATGGG	2220

695

ACCTACTTTA TCGAAAATTA CTACCCCTCAA TTGGCTACTA TCAAGCAACC TTTGGAAGAG	2280
ATTGCTTGTC TCACTATTGA TCTTCTCTTG CAAAAGATTG AAGGCAAGGA AGTCGCCACA	2340
ACTGGTTACT TCTTACCAGT TACGCTATTA CCAGGAAAAA GTATTTAAAC ACAAGAAAAC	2400
TCAGACCGAT TCGTCTGAGT TTTTATGATC TTAAATTTTC GAGATAGCGC TGGGCTGTCT	2460
CTAGGTTAAA GGTTTTATCT GAGATGAGGC GCTCTACTAG GGGAGCAACT TCAGATTCAC	2520
TAGCCCCAGC TAGGAGAGCT AGGGATTGG CCTGTAGTTT CATGTGGCCT TGCTGGATGC	2580
CCGTACTTAC CAAGGCTTTG AGGGCTGCAA AATTTTGAGC AAGACCGATG GACACGATAA	2640
TCTGGGCTAA TTCTCTGGCA GAAGGATTC CTAGTAGATC ATGACTGAGA ACTACACGTG	2700
GGTTGAGGCC GATAGAGCCA CCCTTAGTCG CTACAGGCAT GGGCAGGGTC ATCTCACCGA	2760
CCAATCTTC TCTTTCAAGG TCCAGCGTCC AGCAGCTAAG ACCTTGATAG CGTCCATCTC	2820
GACTGGCAAA GGCATGGGCC CCAGCTTCGA TGGCAGGCCA GTCATTACCA GTGGCAATCA	2880
AAATCGCATC AATACCATTA AAAATTCCTT TATTATGACT AGCAGCTCGG TAAGGATCAG	2940
CCTGCGCAAA CTGACTAGCC AACGCAATTT TCTCCGCAAT CTCTCGTCTT TGATCCTTTT	3000
GGCGGCTCAA GTAGCGAAAG GCGATGCGAC AGCTTGCACT CACCAGAGAA TCGGTGCGGT	3060
AGTTGGACAG GATTCCCATG AGACTCTGTC CCTGACTGAG TTCTTCTAAG ACTGGTTTCA	3120
AGGCTTCCAG CATGGTGTG AGCATATTGG CACCCATGGC TTCCTGGGTA TCGACATGAA	3180
TATAACAAC GAGAAAGTCT GGTTGCGCTT TTATCTGCTC GACATGCAGA TCACGCGCCC	3240
CACCTCCACG TTTAACGATA GAAGGATAGG CTTGATTGGC AAGCTCCAAG AGCTCCGCTT	3300
TCTTGCTGGC AATCTTCTCT TCGGCTAGTT TAGGATTAGC AACTTGATAA AGGGCTACCT	3360
GCCCAATCAT CTGTGCTGA TGGACTTGTG CAGTAAACC ACCTGCACGC TTGATGATTT	3420
TGCTGGCATA GCTGGCCGCC GCAACCACAG AGGGTTCTTC TGTCACATAG GGAACGGTGT	3480
ATTCTGACC GTTGACAAGT ACCTCCGGAA CCAGTGAATA AGGCAGAGAA AAAGTTCCCA	3540
CTACATTCTC ACTCAGCTGG TCTGCCACAG TCACGCTCAT CTGTTTCATC TTCTCCAGAC	3600
TAGCTTGTCT CTCAGGACTA AGGAGCGCCT GAGCTTTTAA CAGCTCGAGG CGCTCTTGGT	3660
ATGATTTTTT AGAAAATCCA TTCCAACCTA TCTTCATTAT TTTTCAACCT TGCTATAACG	3720
GCGTTGGTGG TCGAGAATTT CAACCAAGGC AAAATCTTGA TTTTCATAGC CAGCAAACTG	3780
GGCAGAGTTA GTTTCATCCA AGTTTACTTC CTCAAAAAAG ACCTTTTCAT AGTCTGCAAC	3840
GGATAGGGCA GTTCGTTGGT TGAGCTTGTT CAAACGGTCT TTATCCAAAT AAGCTTCATA	3900
TCCTTCAACC AATTCACCAC TGAAGAAGTC AGCCACAGCT CCACTTCCGT AACTATAAAG	3960

696

GCGGATTTTA TCCCCAGCTT TCAAGCTATC TGTATTTTCC AAGAGAGACA AAAGTCCAAG	4020
GAAAAGTGAA CCTGTGTAGA TATCCCCAC CTTTGGACTG TAGAGAATAG ACTGGTCAAA	4080
ATGCTTTTGT AAGAGGTCTT TTTTCTCTTG AGGCAGGCTC TTATCCATGA TTTTTTCAA	4140
GCCTTTTAGC GCTAATTTAG GATAAGGCAA GTGGAACAA ACAGCCGCAA AATCATCCAA	4200
AGTAAGCTGG TAGCGTTTTT GATATTCAAG CCAAGTCGTT TTCAAATAT CCAAGTATTG	4260
TTGGGTAGAA TAGACACCAT TTACATAAGG AGTTGTCGAG TAATTTGGTC GCCAGAAATC	4320
CATGATGTCA CGGGTCTGAG CTACATTGTC ATTATTAAAG GCCATCATGC GTGGATTTTG	4380
TGTAATCAAC ATAGCTACAC TTCCAGCACC TTGAGTTGGT TCTCCTGGAG TTTCAATACC	4440
GTATTTGGCA ATATCACTGG CAATGACCAA GACCTTGGAC TCCGGAGAAT TTTCCACATG	4500
CAATTTGGCA TAATGGAGGG CAGCAGTCGC TCCGTAGCAG GCTTCTTTAA TCTCGAAACT	4560
ACGAGCAAAG GGCTGGATGC CCAGCAAGCC ATGCACAAAG ACGGCCGCAG CCTTACTCTG	4620
GTCAATTCCT GACTCGGTCG CCACAATGAC CATGTCAACT TCTTGTCTTT CTGCTCAGT	4680
TAAAATAGAG TCACTAGCAC TGGCCGCCAA GGTACGATA TCCTCAGTTA GGGCGCAAT	4740
ACTCAATTC TTAGTAAGA GTCCTTTACT TAATTTTCA GGGTCAATTC CCCTCGCTTC	4800
TGCTAAGTCT TGTAATTTCA AGACATATTG ACTGGTCGCA AAACCAATCT TATCAATACC	4860
GATTGTCATA TTTACCTCTG TTTTATCATT CATGTAAAA ATCGTTCTAT ACTATTTTAT	4920
CACAAATGGC AGTAAAAGAG AGAAAAAGA CTTGATTCAC CAAATCAAGC CTCTTATTGG	4980
TCATCATTTT AAAGAATGAT TAGTTGCTAG AGAGTTCACC GATATAAGTA GCTTTATAAG	5040
CTCCATTCAC AGTTATCAGC TCCTGGAGGA TCAAATTTCC TGAGTAAGTC CTCCCCTCT	5100
CATCTACAAA TTTTGTATAA AACTGACTGG TCGGAATTC TCTGACATCC TTATCAAATG	5160
TCTTATCAAG TGTTTTACTA ACCTTCTCAG CAATCAATTG ATGCTCTTGC CATCCACTTT	5220
GAAACTCTGA GCCCGAACTA GAAACCATGA CTGGGATAAA CAACAAGGTC AGTAGATTTA	5280
CAGACAATAA GGAAAGTAGT AGACTTCCTG CAAACTAGA ATCCTAGTTC ATGATTGA7A	5340
ATACCAGCAA TCAAATTCAT TCGTAATCCG AAGCGTTTAC GATGATTTCG ATAGGTTGTT	5400
GAAAACATTT TAAACGTTTT TACTTTGGCA AAGATGTTCT CAACCTTGCT TCTCTCCTTA	5460
GATAGCGCAT GGTTACAGGC TTTATCTTCA GCTGTTAGCG GCTTGAGTTT GCTGGATTTA	5520
CGTGGAGTTT GTGCTTGAGG ATATATCTTC ATGAGCCCTT GATAATCACT GTCAGCCAAG	5580
ATTTTACCAG CTTGTCCGAT ATTTCTGCAA CTCATTTTGA ACAACTTCAT ATCATGACTA	5640
TAGTTCACAG CGATATCCAA AGAAACAATT CTCCCTTGAC TTGTGACAAT CGCTTGAGCC	5700
TTCATAGCGT GAAATTTCTT TTTACCAGAA TCATTCGCTA ATTGTTTTTT AGGGCGAATTG	5760



697

ATTTTACTT	CCGTCACATC	AATCATTATC	GTGTCCTCAA	AGCTGAGAGG	AGTTCTTGAA	5820
ATCGTAACAC	CACTTTGAAC	AAGAGTTACT	TCAACCCATT	GGCTCCGACG	GATTAAGTTG	5880
CTTTCGTGGA	TACCAAAATC	AGCCGCAATT	TCTTCATAAG	TGCGGTATTC	TCGCACATAT	5940
AGAAAGCGTT	ATCAATTTAT	TTATCTCATT	TTTCAGAAAA	TTCTTTTATT	TCTGTAAAGT	6000
CTACGATACT	CGATGTGTTT	TTATATAATG	ATAGAGTCTG	AGAATCACTG	TTCCGCTAGC	6060
CATTCCAATA	GAGATTACCA	AAGCCAACAT	GACAACCAAG	GTCGCACTTG	CCAGTGCTTT	6120
ATTATAGTCC	CCTGTCACAA	AAAAGGCAGT	TGTTCCGGTAG	GAGAGATAAC	CTGGAACCAG	6180
CGGTGCCAAA	ATGGCCAAGA	TAAAGACCAC	AGCAGGTGTC	TTATAAAGAA	TACTTAAAT	6240
CTGGCTGACA	CAAGAACCAA	TAATGGCTGC	AATGAAGGTA	GCTACAATGA	CATTGGTCGG	6300
TTCTTGAGC	AAGAGATAGA	TTAGCCAGAC	AGTCATGCCC	AAAATCCCTC	CAGGTAAGAG	6360
CATAGACCGT	TGCACATTGA	GTACGATTAA	AAAAGTGATA	ATGGCAAGAA	AACCTGCTAC	6420
TGCTTGTAAT	AAAAAGGTTG	TTAGTGTCAT	ATTAGTTCAT	CAATACCAAG	GCGACAGAAG	6480
TTCTGCCCC	TAAAGCGAGG	GTAATGAGCA	GGGATTCAAA	CATCTTACTC	ATACCAGAGT	6540
TTATGTGGTT	GGTCATAATA	TCACGGACCG	CATTGGTCAA	GGCAATACCT	GGTACAAACG	6600
GCA TGACCGC	ACCAGCTATA	ATCAAATCTG	CCGTTGAAGG	AAAACCTGTG	TAGCGAGCCC	6660
AAAAC TG GGC	AATTATCCCA	AAGACAAAGG	CTCCAGCAAA	GGCTGTCACA	AAGGGAATTC	6720
GGATAAATTT	TTCCACATAG	AGGGAAAAGG	CAAAACCAAA	TAAGGTCGCC	ACTCCTGCCC	6780
CAAGT GCGTC	GTAGATATTT	CCGCTAAACA	TAAGTGAAAA	GAAAGGAGCA	CTAAAGGTCG	6840
CAGCCAGAGT	TACCTGCAAC	TTAGTATAGG	GAAGGGGTTT	GGCTTGCAAG	GCCGTCAATT	6900
GCTTAAAGGC	TGTTTCTAAG	TCAATCTGCC	CCCCAACTAG	CTGACGAGAA	ATCTGGTTCA	6960
CATCGCAGAC	TTTTTCGATG	TTATAAGAAG	AGGAGGTCAC	GCGCTTCATG	CGCAAATATT	7020
GGTATTTTCA	ATAGAGAAAA	AGATAGCGGC	AGGCATGGCA	AGGACATTGC	AATCCACAAT	7080
CCCCTGCGAA	TGCGCGATTG	GAATCATGGT	ATCTTCTACA	CGATGGATTT	CTGAGCCACT	7140
TTTAAGGAGA	ATAGTCCCCG	CTAGCATAAT	CACATCAATG	ACGGCATTTA	ATTCTTTTGA	7200
TTCTTCCATG	CTTTCCTCCT	TTTATCAACT	CCCTCTATTG	TATCACAAAT	CCGGACTCAA	7260
AAAAATCTT	TGCCATGAAA	TCATGACAAA	GATTGATTAC	TCATTTTGAT	TATCCATCTG	7320
CTTTTAAGGA	GTAGCTGAAG	TTGTTTTAGG	TTTGTAGATT	GAAATCTTGA	CTCTAGTCTT	7380
ATTGAGGTCT	ACCTTTTCAC	CTGCTCTAGG	ACTTTGTTCA	ACAACCATGC	CTTCTGCACT	7440
ACCTGCAGGC	GCTGTCGTCA	CTTCTACAAC	TTCTATATTA	GCTTCCTTAA	TCCCAACAAT	7500

698

TTGAATCAAA TTGTTCTTAG TAAACTCCAA GCTAGAACCA ATGTAACTCG GCATGGCAAC	7560
ACTTGTAAC TTTTAGCTA CTGTCAAGAC AATTTGAGTA GGTTTACTCA CATCATAAGT	7620
CGTTCCGGCA CCTGGACTTT GTTTCATAAT CGTTCCTGGT TCGCTTTCGC TGGACTCTTC	7680
TTCTCTATC TTAATCAAAT TCTCAGGAAC CTCTCTCTGC TTGAGTTCTG AGATTACTTC	7740
TGTAGAGTTC CGTCCAATAT AGTTCCTAA TTGAATCGTC GTAGCTTTT TAGCTACTGT	7800
CAAAACAATT TGAGTTGCCT TGCTCAAGTC ATAGGTCGTA CCTTCTGGTA GACTTTGCTT	7860
CAGGACCGTT CCAGCCTCAC TCTCATTCGA CTCTCTTCC TCAATTTTAA TCAAATTATC	7920
TGGAACCTTT TTCTCTTTTA ATTCCGCAAT GACATCAGAG GATTTCCGAC CGACATAATT	7980
ACTAATTTGG AAAGATTGCT TGCCTGATGA GACAACCAA TTGATTTTCG TTCTTCTTT	8040
TCGACCAGTT CCAGCGCCAG GATCTGTACG GATAATCCGC CCTTCTTCCA CCTTTTCACT	8100
AGCCTCTGTC TTCTCTCAC CAATCTCAA ATTGGCTTTT TTGAGCGTTG CCTTGGCCTC	8160
TGCAACTGTC TGACCTGCCA CATCTGGAAT GGCAATGGTT GCAGGAGTTC TGGATAGTAT	8220
CCAAATAAGA GAAGCTGCCA CCAATACAAG GCTGGCCAAC AAAATCAGGT AACGCATCTT	8280
AAATCTATGT TTTTTCGGTG CTTGTGGTTG GTAAGTTTCC TCTGTACAG CCTGGCTTGG	8340
GTTTTTGATT GATTGTGTT CTGTTGCGC TTGAACCTTA GGAATAGATG TCAAGGTACT	8400
CTGAGAAACC TTCGGCAAGG TCTTGGTATC TGCCTTGCTC GTTTCATCAA AGATTAACTT	8460
ACTTTCATTT CTACGATTGT AGGACAAGCT ACTAGACAAG TCCACATACA TCTCTGAAAC	8520
CGAGCGGTAG CGATTGGTCA ACTTTTTAGC AGTTGCCTTG ATAATAACAT TTTCTAAAGC	8580
CTGAGGTACA GATGGATTTT CTGCAATAAC GGACGGCAGG GGTTCCTGGA AATGCTGGAG	8640
GGCAATGGTC ACCGCGCTAT CCCCCTCATA AGGGATATGG CCTGTCAGCA TCTCATAGAA	8700
AATAATCCCC ATGGCATAGA TATCACTCTG CACAGTCGCC TTCGAACCAC GCGCCTGCTC	8760
TGGTGACAAG TAATGAACTG AGCCCAACAT CGAGTTAGTC TGGGTCAGAC TTGTCTCTGC	8820
AAAGGCTACA GCAATCCCAA AGTCTGTGAC CTTGGCAGTC CCATCTGGTG TCAAGAGGAT	8880
ATTTTGAGGT TTCAAGTCCC TGTGAACAAT TCCTCGAGTA TGGGCAAGC GCATAGCCAA	8940
GAGAATTTGT CCCATGATAC GGACTGCTTC TTCATTAGAA AGAGGATAAT GTTCCTTGAT	9000
ATAGCGTTTG AGGTCCAGTC CAGCCACATA CTCCATAGCT AGGTACTGTT GACCGTCTTC	9060
CTCGCCAATA TCTGTTATCC GAACGATATG AGGATGGTCT AGATCTGCCA TAGCTCTCGC	9120
TTACGCGTGA AAACGAGCTA CAGCTATCGG GTCCGTCTGG TAGTTGGTCC TCAGAACCTT	9180
CACTGCCACT TCTTCCCCAT CTAAGATTAA GTCTTTGGCT AGGTAGACAT CCGCCATACC	9240
TCCTCGACCA ATCTGTTTGA CAATCCGATA GCGTCCGGCA AAAATCTTGC CGATTGGAT	9300

699

CATTCTGCAT CCTCCTCGTT CATAGAAACA AGGGCAACCG TAATGTTGTC TAAACCTCCT	9160
GCATTGTTAG CAAAACGAAC AAGTGTCTCC GTTTTATCTG CTAAAGGAAT ATCACTGGTT	9420
ACAATATCAC GAATCTCACT GCCTGAAATC ATGTTGGTCA AGCCGTCCT ATTGAGCAAG	9480
AGATAGTCAC CTGACTCAAG GATAACTGTC CCAAATCAG GCTGAATTC ATCTTTTGC	9540
<u>CCAATAGACT</u> GGGTGATAAT ATTTTTTTGC GGATGAGCTT CTGCCTCTTC TGGTGTCAAT	9600
TGACCAGCCT TGAGCAATTC ATTAACCAAG GAATGATCGC TCGTCAACTG ATGGTATTCT	9660
TCTCCACGAA TCAAGCCGAT ACGCGAATCA CCAATATGAG CATAGATAGC CTGATTATCA	9720
ATAATAGCAA GGACTTCCAA AGTAGTCCC ATGCCTCTGT AAGCTTCATC CTGACCAAGC	9780
TGGTGAATCT TTTGATTTTC AATTTCTAGG TAATGGGCGA ACCATTACAG CACTTCATTG	9840
ACTGTATCGA TCTGGGTATC AACCCAGCT ACACCCAGGT CTGTGACCGC CATTCTACTA	9900
GCGATATTCC CTGCGCGATG ACCTCCCATC CCATCAGCTA AAATAATCAT GGTACGTCCA	9960
GCTCTATTGA CATAGTGGTT GACATAGTCT TGGTTATTTG TTCGTTTCTG ACCAACATCT	10020
GTAAATAATG AAATTTCCAT GTGTCAGTTC CTTCTAATC CGATATCTTG CGAAATTGAC	10080
TGATGAAGAA TCCATCACTT CCATACAATT CAGGTGTAAT GAGGATACAG CCGTCTTTCA	10140
TGATATCCTT ACATTCATGT TCTAGTTTTA CCTGCTCGAA CTCGGGATGA CTCTCTAAAA	10200
AGGCCTTAAC GACTTGAAAA TTCTCCTCTG AGACGATAGT GCAGGTGCTA TAAGTTATTA	10260
TACCACCTTT GCCTAGTATT TGACAAACAC TACCTAATAT TTCTAACTGA ATTTCTGCA	10320
AGGACGCGAA ATCTGCCGTT TCTTTATTGT ATTTGATATC TGGTTTTCCG CGCAAAAGAC	10380
CGATTCCTGA ACAAGGAGCA TCCACCAAAA TCTTATCAAA GGAATCCTGG TCAAAAAACT	10440
CATGCACCTT TCTGGCATCC AATTTTGTAG TTTGAACCCG ATCTTCAACT CCCAGACGTT	10500
GGGCATTTTC TTGAATTAAA TCCAACCTGT GGTCTGTACAA GTCCAGAGCA GTAACCTGAC	10560
CTGTCGTAAG ATAAGAGGCT ATATGGGCTG TTTTCCCACC TGGAGCCGCA CAGGCATCAA	10620
GCACTCGCTC ATCACCTTGT AAATCAAGCG TCGGAGCAAC CAGCTGACTG GACTCATCTT	10680
GGATGGTAAT GGCTCCATCC GCAAACAAAT TATGCCCTGC AAAATGCCCC TGCTCCTTAA	10740
CCAGACCACT GGTGCTAAA AGGGAATTAT TCGCCTCCAA CAAGGCTTGG ATTTCTCTT	10800
TTCGACTTAG GTCTGTACA CGAATACTGG CTTTGTTCG CACTAACAGG CTTTCAAAGA	10860
TGGCTTTTGC TCTCTCTCT CCGTATTCTT CTTGAGTTT GGCAACTAGC CAAACTGGGA	10920
GAGAATAGGC AATGGAGTCA CGCTTGT TTCGCTTGAT GCTAGCAATA TCTGGCCAGC	10980
CTTCACGCAA GATACGGCGA AGGACAGCGT TGACCAATTT TCACTGCCT TTTTACGGA	11040

700

GTCTGGCCAA TTCCACTGCT TCATTAACCA CAGCATGATC TCGAATCTTG TCCAAATAGC	11100
GGAGTTGGTA GGCACATCATG AGAAGAAGGA CATAGAGCCA GCTGTCTAAC TGGTCTCTGT	11160
CTTCGATAAA GTGGGATAGG TACCATTCCA GAGTCAGTTT ACGGGCTACC GTTCCATAGA	11220
CCAGCTCGGT CACTAAGCCC TTGTCTGCTG CCAAAAGTTG ACTTCCCTTT AGATGCTTAT	11280
TTAAGGCGAT ATTTGAATAT GCTTGGTTCA CAAAAACATC CTCTAGCACT GCTAGAGCTA	11340
AACTTCTAGC CGTTTCTACT TTAGTCACCA AATCGTTCTC CTACAGTCAA TGTACGTCCA	11400
ACTCCGTTGA GGAAGGAAGC AATGTCCATC TTAGGCTTAC CAGCTGGCTG CACTTGTTTG	11460
AGGGATAGAG CCCCTTCAGC CGTTGCGACA ATCAATTCTT TCTTGCCGAT AGAGAGAATC	11520
TCACCTGGAT TTCCCTGACC TTCTACTGGT AGGGCTTCAT AAATCTTAAA GCGGTGCCCC	11580
TTAAGGAAAG TATGGGCAAC AGGCCAGGGG TTCATTCCAC GAATTTGGTT AAAGAGTTGA	11640
CGATTGGTTT TGTTCAGTC CAGTTTTTCT TCCTCTGGCT TTATATTTGG AGAGAAGGTA	11700
ACCTGACTCG TATCCTGCGG TTCAGGTTTG ATATCAACCAG CAATATAGGC AGGCAGAGTG	11760
TCCAAAAGCA AATCACGACC AACTAGCGCC AATTTTTCAA ACAAGGTGCC AACATTGTCC	11820
TCATCTGTGA TCGGAATGCT GCGACGAGAA ATCATATCTC CTGCATCCAT TTCCTTAACC	11880
ATTTCCATGA TGGTCACACC AGCTTCCTCA TCCCCTTGAA TCAAGGCATA ATGGATAGGC	11940
GCACCACCAC GGTGTCTAGG AAGGAGGGAG GCATGAACCT TGACAGCAAA GTCCATGCTA	12000
TCAAGGAGTT TGCTTGGGAG AAAGTCCCCA AAAGCAGCAG TCACAATTCC ATCTGCTCCT	12060
AGCTTCATAA GATCTTCCAT CTCTGGACTT CCAGATTAAT TTTCAGGTTG GTAGATAGAT	12120
AGTCCTGCTT CCTTGGCAGC CTGCTTGACT GGGGTTTCTT GGATAACTTT TTTACGACCA	12180
ACAGCACGGT CTGGCTGGGT CACAACGGCT AGAATTTCTG AACGGTCATC TGTCAAAAGT	12240
CCTTTTAAGA CTGTTGCTGA AAAGTCGGGG GTCCCCATAA AGATTAGTTT TGTATATCT	12300
TCTCCTTCTT ATAAAAATTG CTGCGGCTCA TGGTCAATCC TGAGACGGAG CTCACTATTT	12360
TCCCGTTCTT GAGTCAAGGC TAAAACCTGG TTGAGGGTCG ACCCCAGCTC ATCTTCTAAA	12420
CGGTATTTAA TTAAATCTG GTAATGATAG AGGTTGTGGG TACGGGCAAT CGGTTTTGGC	12480
GTGGGCCCA GAATGGGACT GGTCTCTGAC AAGCCTGACC GCAAAATGTT CATGACTTCA	12540
TAGGCACGTT TGAAAACCTC TTCTTCTTTC TTGTGAGAAA GGGTAATACC AATCGTGAAA	12600
TAGTAAGGTG GATAGCCGAG TTGTCGTCTG ATTCCCATT CATAGGCATA AAAGCCTTCG	12660
TAATCTTGAT CCTTGGCAAA TCGAATAGCA TAGTGCTGCC GATTGTAGGA CTGTATCAAG	12720
ACTTGACCTG CCTTTTCAGC CCGACCTGAT CGACCTGCCA CCTGAGTCAA GAGCTGGAAG	12780
GTTCTCTCAG AAGAACGGAA ATCAGGCAGA TTCAAGGCCG TATCCGCATT TAGAACTCCG	12840

701

ACTAGGGTAA CATTGGGAAA ATCCAAACCC TTTGCAATCA TCTGAGTACC AAGTAAAATA	12900
TCCGCTTCCC CTCGCCAAA CTGGTCAAGC AAGGCTTGGT GACTGCCTTT CTTTCGAGTC	12960
GTATCCACAT CCATCCTCAG AATGCGAGCT TGGGGAAAGA GTTCTGCTAG CTCATCATAA	13020
GCCTTCTGAG TTCCCGTCCC ATAGTAACGA ATACTGCGGC TCTTACAGTT AGGACAGACC	13080
TGAGGAATAT CCTTCGAGAA ACCACAATAA TGGCAGTTCA TAGTCTTGGT ATCCATATGC	13140
AAGGTCAGAG AAATATCGCA GTTGGGACAA GTATCCACCG TCCCACACTC CCGACACATG	13200
ACAAAGCTAG AATAACCACG GCCATTGAGC ATGAGAACCA CCTGCTCTTT TTTAACCAGA	13260
CGGTCTTGGG TAGCCTCTAG CAAAGGAGGC GTAAAGTTTG ACGTCTCATT TTGTCCGATA	13320
TAGTCTCGAA AGTCAATCAC TTGAACCTCA GGGATTGTAG CCAAAGGATT GGCACGTTGG	13380
GTTAGACGTA AGTGTGATA GACGCCTTTG CCAGCACGTG CCCGGCTCTC TAAGCTCGGC	13440
GTTGCAGATC CAAGTACCAG AGTTGCTTGA TTATACTGAG CCCGTAAAAT AGCTACCTCT	13500
CTGGCATGGT AACGGGGATT GCTGTCTGCT TTATAAGCCG CTTTCATGCTC TTCATCAATA	13560
ATCATGACAC CCAGATTTTT CAGAGGAGCA AAGATAGCAG ATCTGGCACC AACAACAAT	13620
TGGGCATCGC CACGCTCCAC CTTGCCCAT TCATCATACT TTTCAACATT GGATAATCCT	13680
GAGTGAAGAA TGGCTACCTT GTCCCCAAA CGTGCTATAA AACGCTCGGT CATCTGAGGA	13740
GTCAAGGAAA TCTCAGGTAC CAGCAAAATA GCTGTCTTGC CCTTATCCAG GGCACCTTGG	13800
ATAATCTGCA AGTAAACCTC GGTCTTCCCA CTTCTGTAA TCCCTTGAAG TAGAAAGGGA	13860
GGTTGAGAAC TGCCAATAGA ACTCACAACC GCATCACGGC CCTGTCTTTG TTCTGGATTT	13920
AACTCCAAAG GTCTACTTGC TTCAATTCTT TCAAAATAAG CAGCCGAGCG TTGAACTTCC	13980
TTTTGGACTA TGGTAACAGC ACCTTGATCC ACAAAGAAGT TGACTTGCTC TCGCGAGTAG	14040
GACTCTAACA AGCTAGCCAA GGAAGCGCTC TCTGGATGAG ACAGCAGATA ATCTCTCAGT	14100
TCCAACTTTT TCTTGGCAGG TGTAGAAATC TCAACACCTT GTAATTGAGC ATGGTCAACC	14160
TCATACCAAG ACTGGGTCTT GACCTTCTTT TGATCGACTC CCGTATATTC CAGACCAAGC	14220
AGGCCTTTTC TAGTCAAACG CATCATTTCA GCTTGCTTGG CAAGGTCTAG TGAAGAAAAG	14280
GCTAGCGAAT CTTCTGAACC AAACAGGCGC ACTCGTTCTT CCTGACTCAA GCCTTCCAGA	14340
GGATAGAGAA TCTTGTCTATA GCTAGAATTC AGAAACCTG GAAGCATGGC CTTGAGGATA	14400
GAGATTTTGT AGGAGAAGAC AGATTTGCGT AACTCCTCAG CCAGCCAGAG TTGTTCTGGC	14460
GTGAGAACAG GAGAAAAATC CAGCACCTCT GCAATATCTT TTAATCTTG CTCCATCTCT	14520
TCTCCATCTG ATTGGGACTT CAAACCAAGA ACAATCCCTT GAATCAGGCG ATTACCTTA	14580

702

CCAAAAGGCA CATGAACCCG CATCCCAACT TCCAGCATTC CCTCAAATTC CTCCGGAATC	14640
CTGTAACAT AGGGCTGGTC CGTCTGCATC AAGGGCACAT CTACGATAAT CTTAGCTAGG	14700
GCCATCTTCT CACCTCCTCC TTGTCAGTAC ATTCTTGCAA TAGAAAAAAT AAGATTGAGT	14760
CCCCCAACC TTAAATTTTT TCACCATCTT CTTTTTCTTT AGCAATTTGC TCTTTGATTT	14820
TCTTTTCTTC TTCTTCTTTG CGGCGTTTTT CTTCCTCGAT ACGGCGACGC ACTGCTTCAC	14880
GTTTTCCTTC TGGATCTGGG TGAATTGTAA CGTTTCCTGA TTCGATTTCT TCTAAAGCGC	14940
GAAGAGTTGA TTTTTCAGAC TTGAAACCTT GAGTTGCTGG GGCACCTGCT TCCAATTCGT	15000
GGGCACGTTT TGCTTCCAAG ATTACGAGTG AATATTTTGA AGGAACCTTG TCGAGCAAGG	15060
TATCAATAGA GGGTTTAAAC ATCATTTGCT TGTACCTATT TTCTAAATTT TATCGGGTAG	15120
TTGGAGATTT TGGTAACATC TCCTGATAGT GACCAATGAC ACGATCCACA CAGAAGTGTT	15180
CTGCTTCAAT CACACATTTG ACACGTTTCA CAGCTAGGGG TACCTGATCG TTGACAATCG	15240
CATAATCATA CTCACGCATG AGGGCAATTT CTTCTTGGC CTTTTCGATT CGTTGGGCAA	15300
TCACTTCTGC ACTATCTGTT CCACGACCTA CCAAGCGATC TTGCAATTCA TCCAAATCTG	15360
GTGGTCTCAG GAAGATAAAG ACAGCATCTG GAACCTTTTT CTTGACCTGA AGAGCACCCT	15420
GAACCTCAAT TTCAAGGAAA ACATCGATTC CCTTGTCCAA GGTTCATTG ACATAGGTCA	15480
GAGGAGTTCC ATAGTAGTTA CCGACATATT CTGCGTATTC CAACATCTGT CCTTGACGAA	15540
TCAGCTCTTC AAATCTTCA CGAGTACGGA AGAAATAGTC AACACCGTCC ACTTCTCCAG	15600
GACGTTGTGC GCGTGTCTGC ATCGATACAG AATATTGAAA TTGGTTTTCA GAACTCTCAA	15660
AAATCTCTCT TCTAACGTT CTTTTTCCAA CCCCTGAAGG ACCAGAAAAA ACGATTAGTA	15720
AGCCTCGGTC TGCCATTGTG TCTCCTTTTA GTCAATCTGT GAAATAACAT TTCTCTAGAA	15780
TAATGGCAA AAGCCAGATT ATCCTTTACA GTCTTTCTAT CTAGTGTAAC AAAAAAGCAG	15840
TAATTTTTCA ACTGCTCTTT CTTATTTATT TAGCATAATC TACTGCACGA AGCTCGCGAA	15900
TCACGGTTAC CTTGATATTT CCTGGATAAT CGAGATTGTT TTCAATTTTC TTACGAACTT	15960
TGTGAGCCAA GATTGTGACT TTGTCGTCCT TGATTTTTC TGGATTGACC ATGATACGAA	16020
TTTCACGTCC TGCTTGAAGG GCAAAGCTAG TTTGCACTCC TTCAAAGCCG TTAGCAATTT	16080
CTTCCAAATC ATGGAGACGC TTGATGTAGC TTTCAAGAGA CTCACTACGA GCACCTGGAC	16140
GGGCTGCGCT CAAGGCATCT GCTGCAGCGA CGATAACTGC TATCACGCTC TCAGCTTCAA	16200
CATCTCGTG GTGACTAGCA ATCGTATTCA CCACAACCTG GGGTTCCTTG TACTTACGGG	16260
CCAATTCAT ACCGATTTCA ACGTGGCTAC CTTCAACCTC ATGGTCAATG GCTTTCCCGA	16320
TATCGTGAAG GAATCCAGCA CGACGGGCAA GAGCCGCATT TTCACCAAGT TCGCTCGCCA	16380

703

TGATACCAGC	CAACTTAGCA	ACCTCAATCG	AATGGCGCAA	AACATTTTGT	CCATATGAAG	16440
TACGGAACTG	CAAACGTCCC	ATAATCTTCA	TCAAGTCTGG	ATGAAGGTTT	GGCGCACCAA	16500
TTTCATAGGC	AGCAGCCTCA	CCGTATTAC	GAATCTTATT	GTCAATCTCT	TGACGGTTTT	16560
TCTCAACCAA	CTCTTCGATA	CGAGCTGGAT	GTATACGACC	ATCTTTGAGC	AACATTTCCA	16620
TAGTCATACG	GGCAATCTCA	CGACGAATCG	GATCAAATCC	TGACAAGGTC	ACCACTTCTG	16680
GTGTATCGTC	GATAATCACA	TCGACCCCTG	TCAAACCTTC	AAAGGTACGA	ATGTTACGAC	16740
CTTCACGACC	AATAATGCGT	CCCTTCATAG	TATCGTCTGG	CAGATGAACT	GTTGAGTTTG	16800
TTGACTCCGC	TACATATTCA	CCAGCGATAC	GTTGCATAGC	TTGAACCAAG	ATGTCCTTGG	16860
CCATTTTGTC	AGAACGTTCC	TTGACCTCTT	GCTCAGCTTC	GCGAATGCGA	CTUGCAATCT	16920
CCCTGGTCAA	GTTTTCTCT	GTCTGAGCCA	AGATAATATC	TCGTGCTTCT	GCCTGAGACA	16980
GCGCACCAAT	ACGCTCTAGT	TCTGCTTCTT	TTTGTCTTTC	GACTTCCTCT	AATTGCTCTT	17040
CACGCGCATC	AAGGTTTTTC	GCTCTATCAG	AAATACTTTG	TTCTTTTTGT	TCAAGTGTTT	17100
GTTCTTTACT	CGTCAAATTG	TCGTCCTTAC	GGTCAAGGCT	AGTAGCTCTC	TCTGTCAAAC	17160
GACTTTTCGAT	TTGTTTGAGT	TCTTGACGTT	CTGATTTGAA	TTCAGCGTCC	ACTTCTTCAC	17220
GGTATTTTCT	GGCTTCTTCT	TTGGCCTCCA	ATAGTGCTTC	TTTTTTAAGA	GACTTGCTTT	17280
CACGTTTGGC	TTCATTAACA	AGTAAATCCG	CTTCACGCTC	AGCTTGTCCT	CGTAAATTAG	17340
TTGCTTCTTG	TTCAGCATTT	AAAAGCATCA	ACTCTGCAGC	TTCCTGAGAT	GATTTTCATCT	17400
TAGCTGAGAT	GCTGACATAT	CCAATGACTA	AACCAATGAT	GACGGCAAAA	ACAGCAATCG	17460
CAAGCGACAT	GATTTCCATG	TTTTTACCTC	ATTTTATTGT	TATCCGAAT	GACATACATT	17520
CTTTTACATT	CTACCATAAA	AAAGTGATTT	TCACAAACCT	AAAATAGAAT	ATGTTTTGAG	17580
GAATTTGGAA	CACATTTACC	AAAATAAACT	TGTTGTTTAG	AAATAGTAGT	TTAGTAGAGA	17640
CTTGAGAAAA	AGCCTACCTT	TCAATAGACT	TAGTAATGAT	CTTTAAAGGA	CAAGAAAGCC	17700
ACGCTATCTC	CATCCATCAT	ATAAATCAAG	CGATTTTCTG	CATCAATACG	CCGTGACCAG	17760
GCTCCTTGGT	AATCATATTT	GAGTGGTTCT	GGTTTACCTA	TTCCTGTAAA	GGGATCACGT	17820
TGAATATCCT	TGATTAGTTT	ATTGATTCTT	TTTAACGTTT	TCTTATCCTG	ATTTTGCCAG	17880
TAGCAATAAT	CTGCCCAGGC	ATCTTCTGTA	AACCTTGAGCA	GCATTTCTTA	CTCCTCAATA	17940
ACATGGACCT	GAGTACTTCC	AGCACGAACT	TGAGCCATTG	CTCGCAAAAC	CTTATCAGAA	18000
AGTTCCCTTAT	TTTGAGCAAT	TCTCAGGGTT	TCTTGATAC	TATCCCACTC	ACTCTTTGAA	18060
AGGACTACAA	TGTCCTCATC	TGGATTTTTA	TTGACCACCG	TCAAAGGCTC	AAATTCATCA	18120

704

TTTACCTTCT TCATGTAGTC CTTTAAATGA TTTCGGAATG TTGAGTAAAG GACTGCTTCC	18180
ATAACCATAC CTCGTTTTAG CTCTTTTCCA CTATTATACA CGAAAAGAAA GAAATTGTCA	18240
GGAAGTTGTA CAAGATTTTC TTTTCTATCT ATTTATACTC AATGAAAATC AAAGAGCAAA	18300
CTAGGAAACT AGCCGCAGGC TGTACTTGAG TACGGCAAGG CGACGTTGAC GCGATTTGAA	18360
TTTGATTTTC GAAGAGTATT ATTCGTAAAA AATCTCAAAA AGCCTACCTT TCGGTAGACT	18420
TAGTTTGTTT CTATTC	18436

(2) INFORMATION FOR SEQ ID NO: 88:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7001 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

ACGTAGAAAA ACTATTTCTA TCACAGATAA TATTCCGTAT GTTGTGGAG GTATTGAAAT	60
AAACGTCCTA GGTATCTTTC TCAGTCTATG TGAATTACAA GGGAAAACTC TTTTCGAGAC	120
AGAAATTTTG AATGAAGATT ATCCTATTTT AGAAATCAAT TCCACCATTA CCAATATGAT	180
AAAAACAGCT ATAGAGTACG TCCCTTTGGA AACAAAATTA CTTGGATTGG GCTTATCAAT	240
ACCTGGACAT TATAACAAAG ACTCCGGAAG TATCATTACA AACAAACCCA TATGGGAATC	300
TTTAAATTTA TTAAATGTAA TAAAAGATT CAATTTTCCT TTTATTGTAA AAAATAATAT	360
CGATTGTATG GCTATAGGAC AATACCTTTT TAATCCACAC AATACCCCGG ATAACTTTAT	420
TTTCCTACAC GCTGGATTAG GTATTTACAC TTCCTTTTTC ACAAAGAAA AAATAGGAGC	480
CTCTAAAAAT CCTTATATCG GAGAAATTGG ACACACCATT GTCGAATTGA ATGGGCAATA	540
TTGTGAATGC GGAAAAAAG GTTGTTTACA AACATATATT TCGGATGCTT GGTAAATCAA	600
ACACGCCCAA TTATTATTTA AAAATTCCCA ACTAACTGTA CTAAAAAGCC TTGTAAAGAC	660
TGAAAAAGAC ATTCATTTAG ACACCTTTT AACGGCTTAT AATTTAGGCG ACTCCGCTTT	720
ACGTCAACAA ATTGATAAAG GAGTCAATTT ATTAGCCACT TCTATTGCAA ATCTCCTCCT	780
CATCAATCCT GCTGATAAAA TCTATATCAA CAGTCAATTG CTTAATTATC AACCTTTCAC	840
TCATGAAGTC AGGGATAAAA TCCAAGACCA GCTCCACTTC GTTCCCTTTA CTCGTAATAT	900
AGAAATTGAA ATTTTACCTT ACAACAAACA TCGTGAAGT ATAGGAGCTT GTGCATTAGC	960
TATCGTCGCT TTTTTCATAG AACATAGCAA TGTATTACAA GATATTATTT CACCTTAATA	1020
TATTAGAAAT CTATAGACCT GTTTAAATCA ACTATAACCT GTAGTAGATA TCTCGTATTT	1080



705

AGACAATATG AAAACAAGAC GACTTCCATA TAGGAAACCG CCTTCTCGCT ATGTTGAGTG	1140
ATTTATATTA AAATAACTTT TCTTCTAGCT GCATTTTATT ATTATAAAAA CATTCATCAT	1200
AACCCCCAGA ACTTAAATAA CAATTTTAT TCAAGATACA TACTCCTAGA ATAAACTTTA	1260
TATGAAATTC TCATTTTGT TTTTACAATT CTCCTTAGTT AAATCTTGTT TAATATATGT	1320
TTTACATATA GTATTTAGCG CCACATAGTA CTGAAGCTCT TCCAAAAACG GTTATTCCTC	1380
TTTGAATAGG GCGTTATCAC AAGAAAAGCA TCTCCACGTT TCAACTTCAT ATGGCTCAAA	1440
AACAATCAAT TGATGCTAAA ACCTGTACCT AGATGTTTCG GTTCATAAAA CCATGAAACT	1500
GTAAAAGTGG ATGAAATTGA TAGCGATAGT CAAATCAAGA GGCATCATAA CTCTAAAAAG	1560
TCACAATATA TAAGTTCATC CTCGGAAAAA TATCATTCTA ATTGTTGAAA TGCCTACATG	1620
AAAAGAAACG TCAATGCTC ATGAAACAAC GAATACAGGT ATCAAACTA TGACAAAACA	1680
AATCCCTAAA TTTACTAAAG AACTGTCTCA ACTTTACACC TGAAATGGT TGTGTATATA	1740
TAAAGTTACA AAGATGTACG ACCACACTGT TGTAAATCAT AGTGTTCGGG AATATATTAC	1800
TGATAGCATT TCTACAAATA CAAGTAAAGA GAGCGGATGA GATTCAAACG AAATATGTCA	1860
GTGCTTTGGC ATTCCTAGCC TTCATATCAT TTAAAGAATT CTATAGACAA AATTTTTTCC	1920
AATACAGACA CTCGTAACAA CTGCTTCATT TTTCTACCA CATATTTAGG AACAGGATAA	1980
GATACAAGAG TATTAATCCA TAGCTCAGTT CTATACCAAT CTAAGACAAA TAAGCTAAAA	2040
AAACGATTGA TAATAAGCAA ATAGATTCCA AATTTTCTCT ATCTGCTCAT TTTAATAAAC	2100
AATACTAGTG TAACTATCCT TCCAGTCAGA AGCTTGTCAG ATCACACCGA AAATCTTCT	2160
AAAATTTATC TCGTTAGGCA ATCAAGCAAA AACTCGACGA TAGTACAAAC ATTATCATAC	2220
AGGATTGACT TCCTAAATTA TATACTTTAG TAAGGTTTTG GGATAAGAAA AAAGGTTTCT	2280
TTTACATTTC TAAACATTCT TTTCTAAGAT GAAAAACAGA ATTTTTCGAT TGTGATTTAA	2340
AGCAACAAGA AGATTTTCAG TATCATCCTA TAGATACGAG CTAATTAAGA AAACTACAT	2400
TTTGAATAT AAATACAAT AATATAAACT AAATTTTATA GGAGGAAGAC AATGGATTGG	2460
TACGATTATA TGATACAGGC ATCCAAACAA TCACAATTCA ACGCAAGCCA TTGGTTTCGC	2520
TATTTGCGAA AAGTTATTTT TGAAGACTAT TCTTATTTAA CAAACCAAGA TGTAGAAAAG	2580
TTGCTAGACT CCAAGAACT AACCCGTTTT CAAAAAATTA GCTTGAAGTA TGCCTTTCAA	2640
GAGCATACTC CAACTCATAA ATATGTGATT TCATTAAATA AACCTGCTAA GTTAACCAAT	2700
GTTCAAAAAT TGATGGAGAA ATACAAACAT GGATAAAATG AAACCGGTCT TCCAAGCCCT	2760
AAATAAGGAA TTAATTCAGG AAAATCTGAC TTAAACAATT ATCTGTGTCG GTGGTTATGT	2820

706

CTTAGAATAT CATGGTTTAC GTGCCACACA AGATGTTGAT GCTTTTATGG CTCTATAATA	2880
TTTGTAGTGG GTAAATCCCC TATGGATATT ATGGAGCCTA TTTTGTGTA GAAAAAAGT	2940
CCCATATGAC CTATAATGAA AAGCGACAAA ACAACTCATT AGAAAGAATC ATATGGAACA	3000
ATTACATTTT ATCACAAAAT TACTAGACAT TAAAGACCCT AATATCCAGA TTTTAGACAT	3060
CATCAATAAG GATACACACA AGGAAATCAT CGCCAAACTG GACTACGACG CCCCATCTTG	3120
CCCTGAGTGC GGAAACCAAT TGAAGAAATA TGACTTTCAA AAACCGTCTA AGATCCCTTA	3180
CCTCGAAACA ACTGGTATGC CTTCTAGAAT TCTCCTTAGA AAACGCCGTT TCAAGTGCTA	3240
TCACTGTTC AATATGATGG TCGCTGAAAC TTCTATCGTC AAGAAGAATC ATCAAATTCC	3300
TCGTATTATC AACCAAAAAA TTGCGCAAAA GTTGATTGAG AAGATTCTA TGACCGATAT	3360
TGCTCATCAG CTGGCCATTT CAACTTCAAC TGTCATTGCG AAGCTCAATG ATTCTCACTT	3420
TGAGCATGAT TTTTCGCGTC TTCCTGAGAT TATGTCTGCG GACGTTGAAA CAGTCCGGGG	3480
AGTGACTGTT TCAATCGGGA GATGGAGATG AGCTTTATTG CGCAAGATTT TGAAAAGCTC	3540
GATATCATCA CTGTTCTTGA AGGTAGAACA CAAGCTGTCA TCCGAGATCA CTTTCTTAAA	3600
TATGATAGAG CCGTCCGATG TCGCGTCAAA ATTATTACTA TGGATATGTT TAGTCCTTAC	3660
TATGACTTAG CTAGACAACT TTTCCCGTGT GCTAAATCG TTCTTGATCG CTTTCACATT	3720
GTACAACATC TTAGCCGTC TATGAGTCGT GTCCGTGTCC AAATCATGAA TCAGTTTCAT	3780
CGAAAAATCCC ATGAATACAA GGCTATCAAG CGCTACTGGA AACTCATTCA ACAGGATAGC	3840
CGTAAACTCA GCGATAAACA TTTTATCGC CCTACTTTTC GTATGCATTT AACCAATAAA	3900
GAGATTTTAG ACAAGCTTTT GAGCTATTCA CAAGACTTGA AACATCACTA TCAGCTCTAT	3960
CAACTCTTGC TGTTCACCTT TCAGAATAAG GAACCGGAGA AATTTTTCGA ACTTATCGAG	4020
GACAATCTTA AGCAGGTTCA TCCTATTTTT CAGACTGTCT TTAACCTT CCTCAAAGAT	4080
AAAGAAAAGG TTATCAACGC CCTTCAACTA CACTATTCTA ATGCCAACT GGAAGCGACC	4140
AATAATCTCA TCAAACCTAT CAAGCGCAAT GCCTTTGGTT TTCGAACTT TGAAAACCTC	4200
AAAAACGGA TTTTATCGC TCTGAATATC AAAAAAGAAA GGACAAAATT TGTCTTTCT	4260
CGAGCTTAGC TTTTTCCTT CCACTACAG TTGACAAAGA GCCGAAAAA GGAACAGCCT	4320
TAGCTTTCCT TTCATTTCTT TTTATTTCCC TCGTAGTAAA CGTGCTAGCT TCCACAAAC	4380
AAACAGGATT CCCAGAAATG CCAGTACCAC TAGCCACCG TACAACCATT GAGAGGTTGC	4440
AACACGGAT ACAGATTGTC CTTCTTTCGT AAAAGCAACC CTCGCAACTG CAGCTGTTTG	4500
TGGATCTGAT TTTTGATAAA CAGCGACTCG TTCAAAATC ACTAATAAGC GTTTATTAAA	4560
GGTAGGAATC GCATCGCAGG TTATCAAGGT CATGATATTT TTAGAGCTAA CCGATTCTAA	4620

707

TTTTTCCCAT TCCGACGGTA AAATAATCTC TGTGTCCATC ATCTGATATT CTACAATTTC	4680
CTGGCCATTA TCATAATAAA GAGCATCTCC AACTTTTAGC TGATCCAAAT GGCGGAAAAA	4740
GACATGGCTT GGCTCTGCAC GGTGCCCAGC AATCACTGAG CGAATCCCTG TACCATCCAG	4800
AGGCAGCGGT GTACCATCCA CATGAGCCAA GCCCATCCCT <u>AAATGATGAT</u> AATCTGCTCC	4860
CAAATAAACC GGCTCCATGA TTTCCAAAC TGGAAATAGAC AAGTAACCAT AGACTGCATC	4920
AGGGTCGTCA GACACTTGGT AATTGACCTC ATATCCCTCC GCCAAAAAAG GATCTACAAT	4980
GCGATTTTGC GAAGCCAAGC GTTGATTGTA GGCGAGAGAA TGGTTCTGTT GTTCTTGGA	5040
CATTTAGTT GTCATGGATT TCACAAATGT AGCATGACCT TTCACCTGTC CAAGAGACTG	5100
CAACACCATC TGTCCAAAAC AATAAATAGG AATCAAACAG GCTACCAACA TCAACAAGTA	5160
TCCCAATAAG GCTCGTAGTT TAGTCCTGA CATGACGCCC CTCCAATTGC TTTTCTAGTC	5220
CTTTGACAAT CCGTCGATTA CGATACACGC GATACAGCAA GAGAAGGATG ACCGCCATCG	5280
CTCCTAGTAA TAACCACAAC CAGAATTGCC CACGCTCTCT CACCGCTCGA TTCCGCTCTG	5340
CAATTGGTGC CGTATACGGA ATCCGCTTCC CACGTACCAA CAGACGATGA CTGTTAATCA	5400
TATACGGTGT ACAAGTCAAC AAGGTCGCAT AATCTTCCCC ATGTTGAATC AAGACAGGCT	5460
CAAAGTCATT CGGCTCCACC GTCATATCT GATCCACTTG GTAGGCCAAC ACCTGATCTA	5520
AAACGTGAAG ATAAAAGATA TCCCCTTTT TCATCTTATC CAATTGACTG AACAATTCTG	5580
CCGTGCGCAA TCCTCTGTGA GCAGTGATCA CTGTATGGGT ATTTTCACCT CCAACAGGCA	5640
GCGAAGCCCC TTCTAACAGC CCTGCCCCCT TCTGAAGAAT GTCCTCACTC GTTCCGACAT	5700
ACATCGGAAT TTCCTGATCA ATCGCAGGAA TTCCACATA GCCAATCCGC TCATGGACCT	5760
TTAGCATATT GGCATATTCT GAGACGCCTT TCTTTTCTC TTGCTCTGTA AAAGGATCAA	5820
GAATTTGAGA TGGTTTCAAG GTCGCATTGA AGGCTTGAGC CAAGCGCCAA CGCTCCTCAA	5880
GTTCTGCCTT ATCCATCTGG GAAACCGTCT CATCAAATC TTTAATAACC TCGTTTGACT	5940
CAATACGATA ATAATAACGA GACACCAATG GATATATCGC AACGGCGAAT CCTACTAAGA	6000
AAATCAGAAG AAGGATCAGC GGATGTTTCT TCTTTTTTGT GCCTTTTTTT CGTGAACGTC	6060
TACTGTTGTC CATCTCCAC CTTCACTTCC TTCCTTGCTG CTTTCAGCGC CTTCAAAGCC	6120
TTTTCCGGTT GTTTTTCTT CTTGCGCAAG CGTCGAATAA TCCATAAAAG AATCACAATC	6180
AAACCAACTG CCACATAAAA CAGGTAGCGA TAGAGATGAC TGAGTTTGTT TGCTGCAATA	6240
AATTCTTCCT CAACCTCTGC TACGTACGGT ATCCGATGCC CCCGAACCAA TAGACGATGG	6300
GTATTGATCA TGTATGGCGT ACAAGTCAGC AAGGTCACAT AATCATGACC TGGTACAATC	6360

708

AATAAATCAT CAAAGTTCGT CGGCTCAATC ACCTTTACTT GATCCACTTG ATAGGCCATC	6420
ACTTCCTTGA TATTGTGCAC ATAAAACTTA TCCCAACTT TAAGTTTGGT CAAATCCGTA	6480
AACATCTTAG CTGTTGGCAA ACCTGTATGT GCCGTAATCA CCGCATGGGT CGAATTGCCT	6540
CCGATCGGCA GAGAAGTTCC CTCTAGATGC CCAGCCCCTT GCTGCAATAC CTCTTCAGCA	6600
GTACCAGCAT AAACCGGCAA ATCCACGTCA ATAACGGGGA TTTCCACATG CCCCATCCGC	6660
TCATGGATT CTAACATACG TGCATACTCT GCTCGCCCTT TTTCTTCAT TTCTTCCGAC	6720
CAAGGATCGC CACTCACTAC ATTATTCAA GAGTCATTGA AGGCTTGTGC CAATTTCAAT	6780
CGTTCATCAA TGTCAGCCTC ATCCAACGTT GCTTTTTCCT TATCAAAGTC AGCAATTTGT	6840
TGATTTGATT CCACTCGATA ATACAAGCGA GACACCAGCG GATACGCCAT TACCGCCATT	6900
CCAATGAAAA ATACCACTCC TAATAGGAGA TTATTTCTGT TTTGCTTTTT TGTMTTACC	6960
ATTTTATCA GCATCCCTTT ATCTTCAAAC TTCAGGGTAT C	7001

## (2) INFORMATION FOR SEQ ID NO: 89:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10411 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

GAGGGAGCTT AAGAAGTTAC CACCGTCCTC TAGCGCCTTA TCCGCATCAA AGTTAAGGTT	60
GATATTTTAA AAAGTGTGCG CAGCTTGTGA TACGATGCTT TGTTTAAGGT CATTTAGGGT	120
TTAGTGAAA TCTGCATTGC TGAGGATATC ACTCTTTGAG AGATTCAAGG CAAAATTGAT	180
GATGATATTG ATCTGGTTTC CTGTTATGAC CTGATCAAGT TTGTAATTTT TTAAGGTATC	240
TTCAACAATC TTGCGGATAT CTTCTCTGT CAGATTTCCC TTACTTTCTT TAGCTTTGGC	300
GAGTCCTGAC TTGATATCAG CTAGGGCAAC GTTTAATTTA TTAGCATCAT AGCCTGATTT	360
GTCTTTGTTT TCAGCATTGA TATCTGACAA AGCTTTTAGC TCTTCTTGAG CCAAATCTTT	420
ATTAGCTTGT GGCACCTTGG CTCCATTAGC CTCTAGCGAA TAGTAAATCC CTGCTAAAGC	480
ACTTTCTCCT GTAAGTGGAA TAGGGGCTGC TACAGTGATT TTGGCATGTT CCATACCCAG	540
CGTTACTGCT GCGTTTCGGT ACATATCCTG AGTCACCTTA GTGATATTTT CTGGTGTTC	600
AATCTTGACC TCAAGTGGCG ATTTGTCAAC TAGCTTTTGA ATCTTGGCTG ATGAATACAA	660
CTGTAAGCTA GAGTCATTGG CCACATTCAT GATTTTAGAA TAAACATCAG GTGTCATGGT	720
CTTGAGTTCT TTGGTATCTG TTGAGGCATT GTAGCCCAGT TTTTAAAGAG TTTGATTTT	780

709

TTGGTCTTCA GATAGGGAGG AACCTAGGAC ATATTTCAGGT TGGACATAGG TTTCATCGAT	840
AACTTTTTGA ACATCTGTTG CTGCATGGAC GCTATTTCATA GCTGTTACTG CCCACAAGAT	900
CGCAGCGCTA GTCAGAAAGA GTTCTTTCT CATAGGGAAT TTCCTCCTTT ACTTCTTTAG	960
AGTAATATAT CTATCTTAAA GAAAACCTAT AACAAAAACA CCTGGTCTAG CCAGATGTTG	1020
AAAAGAGAGT GAAACATTTG ATGATGTAAA GGTAAAGTCG TACCTGTCTA GAATAATAAT	1080
AGTTTCCTCC ATTTACATAG AGTTCAGCAC CGTGAAAAAT GGAAATGGGG TGAATATAAC	1140
TATAAGTCTT TCCAGTCCTA TTACCAAGCA AGGGGGCAAC AGTCTCACGA GAGTACTGTT	1200
TGGCTAGAGC CAGGGTATTT TCCTTGCCAT TTTGGGCGAT AAAATCGATA TAGGCAGGTC	1260
CAAATTTATA GGCTTGAACA GCTGTCCAGA TATCTACCCC CTTCTTCTGC GCCAGATAGA	1320
GATTGCCTGT CAGAGTTTGA ATGCCTTGCC GAATGCTAGA GGCATTATCA TTGATGGTGT	1380
TGGTGGAACC ACTTGCAGAC TCACTAGACT GCATAACATC GCCTTCTTTT CCTTTTGT	1440
CAGTATAAAT CATAGCAAGC ACAAGCTCTT CGTTTGCTGG GGTGTCTTGT TCACTCAATA	1500
TTTCTCGCAC CATGGGTGA TAGGTCATGA CTGTTTGAC ATCTTGATGA ACGCGGTAAG	1560
CTTTATAGCC AGCAAAAAGG AAGACTGCTA GTACAAGCAC TCTTCGAATT CGTTTAAACA	1620
TTATTTACTT TGGATATCCT CGATATTTT GATTAAGATA GAGTAGGTTT CATTTTCGTT	1680
TTGGATAAAC TCAACAGACT CGGCGTCTTG ATAGACGTTA TTGGGAACGA TGAGCTCAAT	1740
TCCATTTGAT AAGGAGAGTT TTTGGTTTTC AAATTTCTTT AATTGGCGAC TGGCATCAAT	1800
TTTCATCAAAT TGAACAGGTT CTGGTACGGC TTCTTTGACT TGGTCAATAA AGCTCAAACG	1860
AGCCGTCAGA TTGTTGTCAA AAAGGTCATT AGCCAATTTT TCAGGTGACA ATTCATTGCT	1920
TTCTTCTAGG TTGTTGAAAA TAGCTGATTT GACCTTGAT TGAAATTGAA AATCATCTGT	1980
GTTAAAAGAT TTAGCAATTC TCTGGGCTGT TTTTCCAGT TCCTTGATAG ATTTTTTAGG	2040
AGAAATCTTA GGAGCGACAG CAAGAAGATT ATCTGAAAA TAGTTCAAAA AAGTCCCGTT	2100
GTACTTGATT CGTTTTTCAA TCAGGTGATA CTTCCTACTC TGAAGATTGA CCACCAAGGC	2160
CTCATCAGCT CCTGTTCCAA ATCCAGGCAG GTTATTCTGA GTTAGCTTGA TTGGATTATC	2220
AACTTCTCCT CCGAGGTGGG TCAAGGTCTC CCGCAGGGCA ATTGCAAGA AAGCGAAATG	2280
TTCTACACCT TCTTTAGAAA ATTGCACAAA AATCAAGTCA TTGGTCTTGA GATTTTCAGA	2340
AATGCTAAAC TCCTCTTTCC AGAGATTAGC CAGCGTACT GATGTCTCCA ACAAATCGTC	2400
TGTAATATGA TTGAAGAAGG GATTTTCTTC TTCGAAAATC CCAGTCTTGG CTTTCATCTGA	2460
ATACACATGT TCAATTTTTT TACGCAGGTA TTCTTCGATT TTTGGAGTAA TATTGAGAAA	2520

710

CTTATCTGCT AAGAACAGTT CGGTATCATC CGGACTGAAC TGGTGAATAA TGGCTTTCTT	2580
AATATAAATG TCCATAAAAG TTTTAGTCCT CGTATAATGG GAAGGCATCT GTCAATTCTT	2640
TGACTGCACT TCTCACTTCT TCTAATACAG CCTCATTTTC TGAATCTTA AGGGTTTAA	2700
TGATGAGTTC AGCCACTTTG CGACTTTCTT CTTACCAAAA TCCACGTGCA GTAATGGCTG	2760
CTGCTCCGAT ACGAATCCCA CTTGTCTTGA ATGGTGACAA GCTTTCGTAA GGGATTGAGT	2820
TTTTATTTAA GGTAAATATTG ACTTCATCCA ACAAGTTTGG AGCAACTTTG CCGTTTCTA	2880
CAACTTTAGT CACATCAACA AGGAAGAGAT GGTTTTCAGT TCCACCTGAA ATAATACGGA	2940
AATCAGGGTC TTGCAAGAAG ACATCTGCCA TAGCCTTGCT GTTCTTAATT ACATTGGCAG	3000
CATATTCCTT GAAGGCTGGA TCCAAAACCTT CTTTGAAGGA AACTGCCTTA GCCGCCACAA	3060
CATGCTCTAA AGGACCGCCC TGAATACCTG GGGAAATAGC TGAATTGATT TTTTACAA	3120
GTTCTTCGTC ATTGGTCAAA ATCAAACCAC CACGAGGTCC ACGAAGGGTT TTGTGGGTCG	3180
TTGTTGTTGT GATATGAGCG TATGGAAGTG GGCTTGGATG AAGGCCAGCC GCAACCAAGC	3240
CAGCGATATG GGCCATGTCC ACCATGAGCT TCGCACCGAC AGCATCTGCG ATTTACGGA	3300
ATTTTGAAAA ATCGATAATT TGAGAATAGG CTGAAGCACC AGCTACAATC AGTTTTGGTT	3360
TTACTTCTTG GGCTTGTTTC AAGATAGCAT CAAAGTCTAA GAGTTCCGTT TTAGGATCAA	3420
CACTATAAGA AACAAAGTTG TAGGTTTGAC CAGAGAAGCT AACAGGAGCC CCATGAGTCA	3480
AATGACCACC TGATGCCAAA TCCATTCCCA TAACCGTATC ACCTGGCTCA ATCAAGGACA	3540
TGTAAGCCGC ACAGTTAGCT TGGCTTCCTG AATGTGGTTG AACATTGGCA AATTTAGCAC	3600
CGAAAAATTC TTTTGCGCGT TCAATAGCAA GAGTCTCTAC AACGTCTACT ACATCAGTTC	3660
CACCATAATA ACGGCGTCCT GGGTAACCCT CGGCATATTT ATTTGTCAAG ATAGACCCCT	3720
GAGCTGCCAT AACAGCCTTG GAACTACGT TTTCCGAAGC AATTAACCTG ATATTATTTT	3780
GTTGGCGTTC TTCTTCTTTG GCAATAGCAT TCCAGAGATC AGCATCATAT GCTTTAAAT	3840
CATCTTTGTC AAAAAATCATA GGTCTTCTCC TTTATTGTGT GACTAGTCCA TTAGTTTGAT	3900
TTTACAATAA GAAAAATCAA CTAACAGATG CGAATAAACC GTTCTGTCAT TTTATCACA	3960
GTATAGCCAA CTTTTTCATA AAATGCATGA GCACCCAGAC GATGATTGGC AGAATTTAAG	4020
CGGATAAACC CATAACCACA TCTTTTGTCT TCTTCTTCCA ACCCTTGTAG TAAACTTTTA	4080
CCAATACCTT GACCTTGCGC TTGAGGTGAA ACTGCTAAAG CTAAGATATT AAATCCTGCT	4140
TTGGAATAGA GTGATTGCTA AACTTCAGCG TGGACATATC CAAGTAAGAC ATGATTAGCT	4200
GCATCCTCAT AGCCAAGTAG GAAATGATGG GAATCCTGAG ACAGTCTAGC TAGTTGGCTA	4260
GCCGTTTCCT CTGGACTAAA AGTATAACCC AAAGCCTCTT GGTGATGTC ACATATAGCT	4320

711

TTCACATCAG	TTTCTCTTAA	ATCTCTTAGC	ATCTCATTCC	TCCTCAAAAG	AAATCTTTGG	4380
CAACCGAGCA	AGAATATCTT	CTCGCTTAAT	GGCCCCCTGA	CGTAAGATTT	TCACCTTGTC	4440
TCCCCACAAA	TTCAAAATAG	TTGAATCCTG	TCCAGTTAGA	AAAGCATCGT	CTTCCAGACC	4500
CAGAACCTCT	TGGTCAAAAT	CCTCTAGAAT	TTGATTAAAG	GTCACTCCAC	TCGCCTGACC	4560
TGAGATATTG	GCAGACGGCC	CAATCAAGGG	ACCTGTCTCT	CGAATCAAAT	CAAGGGTAAT	4620
GGGATGACTA	GGCATCCGAA	ATCCAACAGT	TGCAAGGCCA	GAATTGACCC	AATAGGGAAC	4680
TCGGTCATTA	GCTTCGAGAA	TAATGGTCAA	GGGACCTGGT	AAAAAGATCT	CTACAAGTTT	4740
TTGAAGATAA	GTTGGCTGAT	TCTTTGAAAA	GTACAAGATG	TCCTCTAAAG	AGGCAACATT	4800
GAGATTGAGC	GCCTTGCTCT	TACGTCGACG	TTTAAGCTGG	TAAACATGGT	CAACTGCTTT	4860
TTCTGTAGC	GCCTTAGCAA	AGAGACCGTA	AACTGTCTCT	GTAGGCAAAA	CGACAGCTCC	4920
ACCATTTTCC	AACTCTTGTC	TAATCCTGTC	CATCATCAAC	GACAACCATC	CTATCTTGAC	4980
CAAATTGGTC	CTTGAGTGTT	CGTACTCGCT	TTTCAGGAAG	ATGTTTCCTA	AAAAGTTCAG	5040
GAACACTTTG	ACCTTGCTTG	TATCCAATTT	CAAGGTAAAT	CTTACCACCA	TCTTTGAGAT	5100
AGTCTTTTGC	ATCTTCCGCA	ATTCTACGGT	AAATAGCTAG	GCCATCCTCA	TCTGCAAAGA	5160
GAGCTAGATG	AGGCTCCGAA	TACAAGACAT	TCAAGCCTAC	CTCTGACTCA	TCTTCACGAG	5220
AGATATAGGG	TGGATTGGAA	ACAATTATAT	CATATTTTTC	AGAAATTTCT	GTAAACACGT	5280
CAGATTTTTT	TAAAAATATT	TGAAGATTTT	GATTTTTTAGC	ATTTTCGCTA	GCTACATCTA	5340
AAGCATCTTG	GGAAATATCT	GCTGCCGTCA	CTGACCAATC	TGGTCTGTTT	TTTGCTAGAG	5400
CGAGAGCAAT	AGCTCCACTA	CCTGTTCCGA	TATCTAGGAC	CATAAGATTT	TTACAGGAT	5460
TTTCAGCCAG	GATAAGCTCC	ACCAACTCCT	CTGTTTCTGG	ACGAGGAATC	AAAACCCGTT	5520
CATCCACCTT	TAAATGCATT	CCATAAAAAT	CTGCCTGTCC	AATGATGTAC	TGAGCTGGCT	5580
TGTGAGCTGC	TAGTTGCTGG	TAAATATCTT	CTACAAATTG	TTTTTCTTCC	TCTGTTGTCA	5640
CCTCCTGCTG	GAGGGCAAAA	ATAAAGTCTG	TAAAGATAG	ATTTTTCAGA	CTACGATAGA	5700
CAAAAGAGAG	GCTTTCGGCT	TCCTCTCCTT	GTCTTATCAA	CTCTTCTTCA	AAATTTGAAA	5760
ATAATTGAGC	TAATTTCAAT	ATTTGTTTAA	TTCTTCTAGT	TTTTGTGTTT	GGTCATAAAG	5820
CACCAAGGCA	TCCACAACCT	CGTCCAATTT	ACCAGACAAA	ATCGTATCTA	GTTTTTGGAG	5880
GGTCAAGCCG	ATACGGTGGT	CTGTGACACG	GTTTTGTGGG	AAGTTATAAG	TTCCGATCCG	5940
TTCTGAACGG	TCACCAGTAC	CGATTGTCTG	CTTACGCTCA	GCGTCTGCT	CATCTTGAGC	6000
AATCTGAGCA	AAGTGGTCAG	CAACACGGGC	ACGGATGATT	TTTATGGCCT	TCTCACGGTT	6060

712

CTTCTGCTGG	GTACGTTCTT	CCTGCATCTC	AACCTTGATA	TTGGTTGGCA	AGTGAACGAT	6120
ACGAACGGCA	GTGCAACCT	TATTGACGTT	CTGTCCACCA	GCACCAGAGG	CGTGATAGAT	6180
GTGACACGA	AGGTCTTTTG	GATCAATGTC	GTATTCAACC	TCTTCAACTT	CTGGCATAAC	6240
AAGAACTGTC	GCTGTCGAAG	TATGAACACG	GCCTTGCGTT	TCTGTCACAG	GAACACGTTG	6300
CACACGGTGG	GCACCTGATT	CATACTTAAG	CTTAGAGTAT	ACAGACTGAC	CTGAAACCAT	6360
AGCAACCACT	TCTTTAAAAC	CACCGACACC	ATTTCATAGAG	GCTTCCATGA	CTTCAAAGCG	6420
CCAACCTTGG	GCTTCCGCAT	ACTTTTGGTA	CATAGTTAGC	AAATCTCCAG	CGAAAAGTGC	6480
CGCTTCGTCT	CCACCAGCTG	CTCCACGGAT	TTCAAGGATG	ATATTCTTGT	CATCGTTTGG	6540
ATCCTTTGGA	AGGAGCAAAA	TTTTCAGTTT	TTCTTCATAT	TCTTCTTTTT	CAGCCTTGGC	6600
ATCTTTGAGT	TCTTGCTTGG	CCAATTCTTC	CAAGTCCGCA	TCTCCGCCTG	ATTCTCTAAT	6660
CATCTCTTCG	GCATCGACGA	TATTTTGAAG	GACTTGTTTA	TACTCACGGT	AGGCTATTAC	6720
GGTGTCACGA	TTGGAAGCTT	CTTCTTTTGA	AAGCTCCATA	AAACGCTTGG	TGTCTGAAAC	6780
GACATCAGGG	TCACTCAGCA	ATTCTCCTAA	TTCTTCATAA	CGGTCTTCTA	CAACTTGTAG	6840
TTGATCATAG	ATGTTCAATT	TTTCTCCTTA	TTTCTCAATT	GTAAATCAT	AGATTGCTAC	6900
TACTTCATTC	TCGATATTT	CCCCAGTTTC	TTTAAATCCA	TAAGTGGGT	AACAAAATCT	6960
TGCCTGTTC	TTTCTGGTT	CATATGACAA	CCAAAGTTTA	TTGCTTAAAC	CTGCTGGCGC	7020
TGTTGGAACA	TAGTCTAGTA	CTTTATCCAT	AATTGGTTTA	AAATATCCTT	GATTTTGAAA	7080
ATTCTTATCA	ATCATAAAAC	GAAATAGTAA	ATAATTTC	CTACTAATTC	CGATCTTTTT	7140
ATCATAAGCT	ATCATCACA	AACCTATAAT	TGCATCATT	TCATAAACTG	CCAATGGAGC	7200
TACAAAATCT	CCATTTTATG	TGTAGACGTA	TGCTTCAGCT	AACTAATTG	CGTTGGTTGC	7260
AATGAATTGT	TTTGATATT	CCTTGACATC	CAAAATTTAA	ACATCAAAAT	AATTTTCCAT	7320
TGTAACATCT	CTTAGTTCAA	TTGTCATAGT	TTTGCTCCTT	GTTAGAGGTT	ATCATTGGCG	7380
CAAAAATAATG	TTTACGGCAA	ACTGAGATAT	AGGTTTCGTT	ACCACCAATC	TGGATCTGTT	7440
CTCCATCGTA	AACGGGCAGT	CCATCCTGTG	TTGCAACAC	CATGGTCGCC	TTTTTCTTGC	7500
AATACTGACA	GATGGTCTTG	ATTTCTGCAA	TCTTGCTGTC	TAAAAGCAAG	AGATATTTGG	7560
AACCTTCGAA	CAATTCAATTG	CGAAAGTCAT	TTTTCAAGCC	AAAAGCCATG	ACGGGTATGT	7620
CTAACTCGTC	CACAACACGA	GCTAGGTCGT	AAACATGGTG	GCGTTTGAGA	AACTGGGCTT	7680
CATCGACCAA	AACACAGTAA	GTTTTTCTG	GATAGTCTCG	GATATAGCCA	AAGATATCCG	7740
TTGTTTCCTC	AATCGCAAAG	GCAGGGCGTT	TCATGCCAAT	TCGACTCGAC	ACATAGCCAA	7800
CGCCGTCACG	CGTATCCAGA	GCCGAGGTCA	TAATCACAAC	ACCTTTTCCT	TGCTCCTCGT	7860



713

AGTTATAGGC CACTTTGAGA ATCTCAATCG TTTTACCAGA GTTCATGGTC CCATAACGAT	7920
AGTACAACTG TGCCATGTTT CTTGCTTCAC GTCCATTCTT AAATTTTTCG TACATTCTAG	7980
TATATCATAA TTTTCTTAAG CTTTAAACGG CAAAATGTGG TAAAATAGAA GAAATCAAAA	8040
ACTAGTGGAG GAAGCTATTA TGCCATTGTG ACGCATCGAT TTATTTGAAG GACGCACGCT	8100
CGAGCAAAAG AAAGCTCTTG CTAAGGAAGT AACGGAAGCA GTTGTCGCA AACTGGAGC	8160
CCCTCAATCT GCTGTCCATG TCATCATCAA CGACATGCCA GAAGGAACTT ACTTCCCACA	8220
AGGGGAAATG CGTACTAAAT AAGCTAGCTT AAGCAGAATT GCTTAGGCTT TTTCAATCTC	8280
CAAGTAGCAT TCATTGAAGA AATATCCTAA ATTTGTTACA ATTTGAAAAG AAAGTTGGAG	8340
AATTTCCAAG AAAAGAGCTA TTAATTAAAG GAAACATTAT GATTACACGT GAATTTGATA	8400
CCATCGCTGC TATCTCTACT CCACTAGGTG AAGGGGCTAT TGGTATTGTC CGCCTGAGCG	8460
GAACAGACAG TTTTGCTATT GCGCAAAAGA TTTTAAAGG AAAAGACTTG AACAAGGTTG	8520
CCAGCCACAC TCTCAACTAC GGTACATTA TTGATCCTCT GACTGGTAAA GTCATGGACG	8580
AGGTTATGGT TGGGGCTATG AAGTCTCCAA AGACCTTCAC TCGTGAGGAT ATTATCGAGA	8640
TTAACACCCA CGGTGGGATT GCGGTGACCA ATGAAATCTT CCAGCTAGCT ATTCGTGAAG	8700
GGGCTCGGTT GGCAGAACCT GGTGAATTTA CCAAACGTGC TTTTTTAAAC GGTGCGGTAG	8760
ACTTGACACA GGCAGAGGCT GTGATGGATA TCATCCGTGC CAAGACTGAC AAGGCCATGA	8820
ACATTGCGGT CAAACAATTA GACGGCTCCC TTTCTGACCT CATTAACAAT ACCCGTCAAG	8880
AAATCCTCAA TACACTTGCC CAAGTTGAGG TCAATATCGA CTATCCTGAG TATGACGATG	8940
TTGAGGAAGC CACTACTGCT GTTGTCCGAG AGAAGACAAT GGAGTTTGAG CAATTACTAA	9000
CCAACTCCT TAGGACAGCA CGTCGTGGTA AAATCCTTCG TGAAGGAATT TCAACGGCTA	9060
TCATTGGACG TCCCAACGTT GGGAAATCAA GCCTTCTCAA CAACCTCTTG CGTGAGGACA	9120
AGGCTATCGT AACAGATATC GCTGGGACAA CACGAGATGT CATCGAAGAG TACGTCAACA	9180
TCAATGGTGT ACCTCTCAA TTGATTGATA CAGCCGGTAT TCGTGAAACG GATGATATCG	9240
TTGAACAAAT TGGAGTTGAG CGTTCGAAAA AAGCTCTTAA GGAAGCTGAC CTAGTTCTGC	9300
TAGTACTAAA CGCTAGTGAA CCACTAACCG CCCAAGATCG CCAACTCCTA GAAATCAGTC	9360
AGGAGACTAA TCGCATTATT CTTCTTAACA AAAGTACCT GCCTGAAACG ATTGAACTT	9420
CGGAATACC TGAAGATGTC ATCCGCATTT CAGTTCTTAA AAATCAAAAC ATCGATAAAA	9480
TCGAAGAGAG AATCAACAAC CTCTTCTTTG AAAATGCTGG TTTGGTTGAG CAAGATGCTA	9540
CCTACTTGTC AAACGCCCGT CACATTTCTT TGATTGAGAA GGCCGTTGAA AGCCTACAAG	9600

714

CTGTAAACCA AGGTCTTGAA CTAGGGATGC CAGTTGACTT GCTTCAAGTT GACTTGACCC	9660
GTACTTGGA AATTCTAGGA GAAATCACTG GAGATGCTGC TCCAGATGAA CTCATCACCC	9720
AACTCTTTAG CCAATTCTGT TTAGGAAAAT AAGAAAAATC CATGATCCTT CATTTCGGTCA	9780
TGGATTTTAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT TATGGAGCCT	9840
ATTTTATTGT AGAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA AACAACTCAT	9900
TAGAAAGAAT CATATGGAAC AATTACATTT TATCACAAAA TTAGTAGACA TTAAAGACCC	9960
TAATATCCAG ATTTTAGACA TCATCAATAA GGATACACAC AAGGAAATCA TCGCCAAACT	10020
GGACTACGAC GCCCCATCTT GCCCTGAGTG CGGAAACCAA TTGAAGAAAT ATGACTTTCA	10080
AAAAACCTTC TAAAATTCCT TATCTTGAAA CGACTGGTAT GCCCCACTAGA ATTCTCCTTA	10140
GAAAGCGTCG ATTCAAGTGC TATCACTGTT CAAAAATGAT GGTCGCTGAA ACTTCTATCG	10200
TCAAGAAGAA TCACCAAATC CCTCGTATCA TCAACCAAAA GATTGCTCAA AAGTTAATTG	10260
AAAAGATTTC TATGACTGAT ATTGCCCCATC AGCTTTCCAT CTCAACTTCA ACTGTTATTC	10320
GTAAGCTCAA TGACTTTCAC TTTAAACATG ATTTTCTTG TCTTCCTGAG ATTATGTCTT	10380
GGGATGAGTA TGCTTTTACA AAAGGAAGA T	10411

(2) INFORMATION FOR SEQ ID NO: 90:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2393 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

GTTTTGGGTT CTGGAAATTA TCAGATGGTT GGAAAGCCG TCCACATCAA GATAGTGTTT	60
GGAGATTTAA GTTTAAATTG AAGAACTAA CACAGAGGAA ATGGAGTATA GACCTAACAA	120
GACGTATTGA GCAACTGAAT TTGTCTATTC GAGGATGGAT AACTATTGC TCATTGGGAA	180
ATATGAAAAG TATAGTCGCC AGCATAGATG AGCGCTTGGC TACTCGCCTA CGAGTGATT	240
TCTGGAAGCA ATGGAAGAAG AAATCGAGAC GATTATGGGG ATTGCTTAAG TTAGGAGTTC	300
CTAAATGGAT AGCAGATAAG GTATCTGGCT GGGCGACCA TTATCAATTA GTAGCTCAGA	360
AGTCGGTACT TAAACGTGCT ATATCAAAAC CAGTCCTGGA AAAACGTGGA CTGGTTTCGT	420
GTTCGGATTA TTACCTTGAA CGACATGCGT TAAAAGTTAG TTGAACCGCC GTATGCCAAA	480
CGGCACGTAC GGTGGTGTGA GAGGGGCTAG AGATTATCCC CTACTCGATT AACTCCCCTG	540
AAATTTATTT TAATTATGCA AATTCACGT ATTTTGTATG CTGAGACGAC GATCCTGGGA	600

715

ACTTTTCAGA TATTTTITG ACTATCTAAA TCTATCATT	GAAAAGCTTA GAGCGCCAAA	660
GGATTGAGC GTTTTCTGA TTTTAAAGAC TTTTCCAGT	CTCTTTTTCG ATTGAAGATG	720
TAATTATTCT ACTAACTAAC TAACTTCTTA GTACTAGCCA	ACAACGATAA TCATAATTCC	780
TCCTAAAAAT AGGAATAATA AAGGCAATAG TTTTGTGTTT	TTCATGTAAA AAACCTCACT	840
TTTGTITTTCT GCTATTTTAT GCTAAAATAT TAAAAATCAA	ATTTAATTCC AAAGTTTGTA	900
ACTAAAGGGG GAGCGCTACA TGTCTAATTC ATTTGTCAAG	TTGTTAGTCT CTCAATTATT	960
TGCAAAATTA GCAGATATTT TCTTTAGAGT AACAACTATT	GCTAACATAT ACATTATTTT	1020
AAAATCAGTA ATTGCCACAT CACTAGTTCC TATCTTAATA	GGAATATCCT CTTTGTGTC	1080
GAGTCITTTA GTTCCGTTGG TTAATAAAAG GTTAGCGCTA	AATAGGGTTT TATCTTTATC	1140
TCAATTTGGA AAGACTATAT TATTGGCGAT ACTGGTAGGA	ATGTTTACCG TAATGCAATC	1200
CGTAGCGCCT TTGGTGACCT ATCTATTTGT TGTGCAATT	TCCATACTAG ATGGTTTTC	1260
AGCACCCGTT TCCTATGCTA TTGTGCCACG CTATGCGACC	GATTTGGGTA AGGCTAATTC	1320
AGCCTTATCA ATGACTGGTG AAGCTGTTCA ATTGATAGGT	TGGGGATTAG GTGGACTCTT	1380
GTTTGCAACA ATTGGTCTGT TACCTACCAC GTGTATCAAT	TTAGTCTTGT ATATCATTTT	1440
TAGCTTTCTG ATGTTATTTT TCCCTAACGC TGAAGTGGAG	GTGTTAGAGT CAGAAACTAA	1500
TCTTGAAATT TTGCTCAAAG GTTGGAGTT AGTTGCTAGA	AATCCTAGAT TAAGACTTTT	1560
TGTATCAGCA AATTTATTGG AAATTTTTC AAATACGATT	TGGGTTTCTT CCATTATACT	1620
TGTTTTGTA ACGGAGTTAT TAAATAAAAC GGAAAGTTAC	TGGGGATATT CTAATACAGC	1680
ATACTCTATT GGTATTATAA TTAGTGGCTT AATTGCTTTT	AGGCTATCTG AAAAGTTCTT	1740
TGCTGCTAAA TGGGAACCCC AATTATTCAC CCCAAATCTA	AAAACCATCC AGAATCCTTG	1800
CCTTAGCTTA GATCCTGGAT GGTTCITTTT TTCACCCAAT	GGGTGTTTTT TACTAGACAA	1860
AAAAGAGTTT CCCCTTTATG GTATAAGTGT AGAAAAAAC	ACAAAAAGAA AGGAAACTCA	1920
CATGAACAGT TTACCAAATC ATCACTTCCA AAACAAGTCT	TTTTACCAAC TATCTTTCCA	1980
TGGAGGTCAT TTAACCCAGT ATGGTGGTCT TATCTTTTTT	CAGGAACTTT TTTCCAGTT	2040
GAAACTAAAA GAGCGATTT CTAAGTATTT AGTAACGAAT	GACCAACGCC GCTACTGTCG	2100
TTATTCGGAT TCAGATATCC TTGTCCAGTT CCTCTTTCAA	CTGTTAACAG GTTATGGAAC	2160
GGACTATGCT TGTAAGAAT TGTCACTGA TGCCTACTTT	CCAAAATTGT TGGAAGGAGG	2220
GCAGCTTGCT TCACAGCCAA CCTTATCCCG TTTTCTTTCC	AGAACTGACG AGGAAACAGT	2280
CCATAGTTTG CGATGCCTCA ACCTTGAATT GGTGGAATTC	TTTTTACAGT TTCACCAGCT	2340

716

AAACCAACTC ATTGTAGATA ACGATTCTAC CCATTTTACA ACTTATGGCA AGC 2393

(2) INFORMATION FOR SEQ ID NO: 91:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4762 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

TTTGTATCTT TTTAGGTCTC TTTCAATCCA AACCCCTTAA ACTATACGTC ATTTCCGGTTC	60
CTGCAAGTCT TGTGGTAATT TTAGGTTTGA TTTTACTTTT CTTTTCACAA GAGCCTCTGC	120
ACGCTTCTTA TTTGATGGTC GTCTCCCTG TTTTCTACT TTTATGGTA ACCAATATTA	180
AGAGTCAACA GAGGGGGCGT AGTGCTAGAA GAAGCCGAAG AGAAACGCCA TTATGCCTAT	240
GGAGTCGTTT CTTCAAAGGA AATCTATATC TGCTAGTTTT TCGGTTTGTC TATCTTTTGT	300
CTGTTCCCTT TTTGATGAAG TTTGTCCTTT ATCCAGTACC TTATCAAGAA CGTAATCGTC	360
TTGCTGATTT GGTAAAAGAG GAGACAAATA CGGAAGATGC TATCTCATGC ATGGGATGAT	420
ACTGCGACTC TTTATCGTAA GAGTGAGCGC TTGTCCCATC GGCGATTTTG TCCCCGTTGC	480
ACTATACAGC AACTGAGGAA AATCGTAATA AGTTACTTAA TGACTTGAAA GAAAAACAAC	540
CTAAGGTGAT TGTGGTAAAT GATAAGGTGG TAGTCTGGTC TGAAGTGAA AACTCTTAA	600
AAGAAAATTA CCAACRACTA AAGACTGATT ACTCAGAGTT TAAAGTCTAT AAAATTAAAT	660
AACCAAATCA ATATCTTG TG TATTTTAAA AATTTTAGGA TTTTAAACAC AAGATATTGA	720
TTTTTCTTT TAGAGTGGTA TAATACTTTT TAGAAAGAAC ATTTTAGAAA AGAGCATGCA	780
TATGATTGCA CTAGAAGAAA AAATTACAAT TTTGCCAACT CTCTTCGTCG AGAAACGAGA	840
TGGGAGACGT GTTGATTTG ATGTGGACAA GATTGACAAG GCTCTCCACA AGGCGGCTGA	900
CAAGGTTATG GATGTGACAC CCCTGGTTGA AAAATGCCTC AATGATCTGA CTGAGCGAAT	960
TATTACAGAA ATTCATAGTC GCTTTCACAA GGGAATTAAG ATTTACGAAA TTCAAAATAT	1020
CGTAGAACAT GAACTCCTTG AAGCCAAAGA ATATGCGCTG GCTCAGGACT ATATTACTTA	1080
TGGGACACAG AGGGATTTTG AGCGCTCAAA AGCGACGGAT ATCAACTTTA GTATTCTATA	1140
ACTTCTCAAC AAAGACCAGA CAGTTGTCAA TGAACCGCT AATAAAGACA GTGATGTCTT	1200
TAACACTCAG CGTGATTTGA CAGCAGGGAT TGTTGGGAAA TCAATCGGAC TGCAAATGCT	1260
TCCTAAGCAC GTAGCCAATG CCCACCAAAA GGGGATATC CACTATCAGC ATTTGGACTA	1320
CAGTCCCTAT ACCCCTATGA CCAACTGCTG TTTGATTGAT TTTAAGGGTA TGTTGGAAAA	1380

717

TGGTTTTAAG ATTGGAAATG CAGAGGTAGA GAGTCCCAAG TCTATCCAGA CTGCGACAGC	1440
ACAGATTTCT CAAATCATTG CCAACGTTGC TTCTAGCCAG TACGGTGGCT GTTCAGCTGA	1500
CCGTATCGAT GAAATTTTGG CGCCTTATGC AGAGAAGAAT TATCAAAAAC ATCTCAAAGA	1560
TGCAGAAGAG TGGGTATTGC CTGAAAAACA GGAAGATTAC GCTTGAAGA AAGCGCAAAA	1620
GGACATCTAC GATGCCATGC AATCTCTTGA GTATGAAATC AATACTCTCT TCACTTCAAA	1680
TGGACAAACA CCTTTTACTT CGTTAGGTTT TGGTCTGGGA ACCAGTCGTT TTGAACGAGA	1740
AATCAAAAA GCTATTTTAA ACATTCGCAT CAAGGGTCTT GGTCAGAAC ACCGTACGGC	1800
TATCTTTCCT AAACCTATCT TTACGCTTAA AAGAGGCCCTC AACTTAGAGG AAGGAACTCC	1860
CAACTATGAC ATCAAGCAGT TGGCTCTAGA GTGTGCAACC AAGCGGATGT ATCCAGACGT	1920
CTTGTCTTAT GATAAGATTG TTGATTTGAC AGGTTCTTTC AAGGTGCCTA TGGGCTGCCG	1980
TTCTTTCCTT CAAGGGTGA AGGATGAAAA TGGTGTAGAA GTCAATTCAG GTCGCATGAA	2040
TCTGGGTGTT GTGACGGTTA ATCTGCCTCG TATTGCTCTT GAGTCTGAAG GTGATATGAA	2100
TAAGTCTGG GAAATCTTCA ACGAGCGAAT GAATATCGCA GAAGATGCTC TTGTTTACCG	2160
TGTCGAACGC ACTAAAGAGG CGACACCAGC GAATGCTCCT ATTCTTTATC AGTACGGTGC	2220
TTTTGGCCAT CGTCTAGGTA AAGAAGAAAG TGTTGACCAG CTCTTTAAGA ATCGTCGTGC	2280
GACCGTTTCG CTGGGCTATA TCGGCTTGTA TGAAGTAGCG ACAGTTTCTT TTGGTAACAG	2340
CTGGGAAAGT AATCCAGATG CTAAGGAATT CACGCTAGAC ATCATTCACG ATATGAAACG	2400
CCGTGTAGAA GAGTGGTCAG ACCAATATGG CTACCATTTC TCTATCTACT CAACACCATC	2460
CGAAAGTCTG ACAGACCGTT TCTGCCGACT AGATATAGAC AAGTTTGGCT CTATTCCTGA	2520
TATCACAGAC AAGGAATACT ACACCAACTC TTTCCACTAC GATGTTGTA AAAATCCAAC	2580
ACCGTTTGAA AAATTGGA CTAGAGAAAGT CTATCCGGAA GCAGGTGCGT CAGGTGTTT	2640
CATCCATTAT TGTGAGTATC CAGTCCTTCA GCAAAATCCA AAGGCCTTGG AAGCTGTCTG	2700
GGATTATGCT TATGACCGTG TAGGCTATCT AGGCACCAAT ACTCCGATTG ACCGTTGCTA	2760
CAAGTGTGAC TTTGAAGGGG ATTTTGAACC AACTGAGAGA GGGTTTGCTT GTCCAAACTG	2820
TGGCAATAGC GACCCTAAAA CAGTAGATGT GGTGAAACGA ACTTGTGGCT ACCTAGGTAA	2880
TCCTCAAGCA AGACCGATGG TCAACGGGCG TCACAAGGAA ATCGCTGCGC GTGTCAAACA	2940
TATGAATGGT TCAACGATTA AAATAGCTGG GCATCAAGTA ACAAATTAGA AAGAAATGAA	3000
ATGGGAAAAT ATCAACTAGA CGATAAGGGG CGCGCACAAAG TGACCCGTTA TCACGAGAAA	3060
CACTCTAAAG GTGGAGCTGG TAAGAAAGAA CGCTTGCTTA GCTTCAGACA ACAATTTTAA	3120

718

AACAAGAACA AGAAAAAATA AAAGTGAGAG CCAGCTCTCG CTTTCTCAT AGTGGGAGGT	3180
AAGGATGGAA TTACGCAGAC CAAGATTAGC GGATAAGAAA GCTGTTTGTAG ATATGATGAC	3240
AGAGTTTGAA AAATTTCACT CGCCTCACGA CGGCGGTTTC TGGGATACAG AGAACTTTGT	3300
GTATGAAGAC TGGTTAGAAA GCAATCAGGA ACAGGAAATG GGGATTAATC TGCCTGAAGG	3360
ATGGGTTTCT GCAATTCAGT TAGTGGCTTT TTCTGAGAAA GGTCAAGCAG TTGGATTCT	3420
TAATCTCCGG TTGCGCCTCA GTAACCTTCT ACTAGAAGAA GGTGGCCACA TTGGCTACTC	3480
CATTCGTCCA TCTGAAAGAG GCAAGGGTTA TGCAAAAGAG ACTCTCCGTC AGGGCTTGCA	3540
AGTTGCTAAG GAAAAGAACA TCAAGAAAGC TCTGGTGACC TGTAGTGTA ATAATCCTGC	3600
TAGCAGAGCA GTCATTCTAG CAAATGGTGG AATATTTGAG GATGCTCGCA ATGGAGTCGA	3660
GCGTTATTGG ATAGAGGTAG CGAATGAATA ATCCAAAACC ACAAGAATGG AAAAGCGAGG	3720
AACCTTAGTCA AGGTCGTATC ATTGACTACA AGGCCTTTAA CTTTGTGGAC GGCGAAGGCG	3780
TGCGCAACTC TCTCTATGTA TCAGGCTGTA TGTTCCTACTG CGAGGGATGT TATAATGTTG	3840
CGACTTGGTC TTTTAATGCT GGCATTCCCT ATACAGCAGA ATTAGAAGAG CAGATTATGG	3900
CAGACCTTGC CCAACCCTAT GTTCAAGGCT TGACTTTGCT GGGAGGGGAG CCTTTTCTCA	3960
ATACTGGGAT TCTCTTGCCA CTTGTTAAGC GGATTCGGAA GGAATTGCCA GACAAGGACA	4020
TCTGGTCCTG GACCGGCTAC ACTTGGAAG AAATGATGTT GGAACTCCA GATAAACTGG	4080
AATCTTGTG ACTGATTGAC ATTCTGTGCG ATGGAAGATA TGATCGAACT AAGAGAAATC	4140
TTATGCTCCA GTTTCGAGGT TCATCTAACC AACGAATTAT CGATGTGCAA AAATCGCTCA	4200
AAAGTGGGCA AGTAGTGATT TGGGACAAGC TCAATGACCG AAAAGAAAGC TATGAACAGG	4260
TGAAGAGAGA ATGAAGAAAA AGGACTTAGT AGACCAACTA GTCTCAGAGA TCGAGACGGG	4320
GAAAGTCAGG AACTGCGGAA TATACGGTCA TGGAGCTTCA GTAAATCAA CCTTTGCACA	4380
GGAATTGTAC CAAGCTTTAG ATTCTACTAC AGTAAATTG CTAGAGACAG ATCCTTATAT	4440
CACCTCAGGA CGCCATCTGG TAGTACCCAA GGACGCGCCG AATCAAAAGG TGACAGCCAG	4500
TCTGCCAGTG GCGCATGAAC TGGAGAGTTT GCAGAGAGAT ATCCTTGCTT GCAGGCGGCT	4560
ATGGATGTCT TGACAATTGA AGAACCTTGG AAGGCTAGTG AGGTCTTGTC TGGAGCCAAA	4620
CCAATTTTGA TTGTCGAAGG GATGTCTGTT GGCTTTCTAC CCAAGGAACT CTTTGAAAAA	4680
ACCATCTGTT TCTACACGGA TGAGGAGACC GAATTAAAGC GACGCCTTGC TAGAGATACG	4740
ACTGTGAGAA ATCGCGATGC GG	4762

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:

719

(A) LENGTH: 3832 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

GATGCAGGTT TCGACCCACA TATTCCAGAA AATTACTTTA AAGATGATGA TGTTAATCAG	60
GTACCTTGTC TTTGTTGGTC TTCATCTGCA GCCCTCTTTT TCAGTAATTG GGTAGACCAT	120
GCGGTCTATC AGGAGACGCC TTTTGATTGG AGAAAGATAG AAGATGATGC ATCTGCATAT	180
GGGTATTTAT AAGAGGAATT ATGACATATT TAGACGCTTT TAAATCAGGT ACCTTGGTTT	240
TACCGAGTGC CCTGCTCTTG CATTTTAAGG AACTCTTTCC TTCTAGCGAC GATTTTCTGG	300
TTTGCCAATT TTTCTATTTG CAAAATACGA CAGGCTTAGA AGAAATGTCG CCAAGCCAGA	360
TTGCTGAAAG GATTGGCAAG GAAATTTCCG ATGTCAACCA GTCCATTTCT AATCTGACCG	420
AAAGGGGACT GCTCCAGTAT CGTACTATCG AATTAAATGG CGAAATTGAA TTGCTCTTTG	480
ATGCTAGTTT GGCCTTGGA CGTTTGATG ACCTGTTTGG AGCAGTTCAT TCAAGTTCAG	540
ACCAGCTAAC ACCTCAAAAC CAGCTCAAGG ATTTGGTGGA AACCTTCCAG CAGGAGTTGG	600
GACGATTGTT GACGCCTTTT GAGATTGAGG ATTTGACCAA GACACTAAAG GAAGATGGAA	660
CCAGTGCTGA CTTGATTAAG GAGGCTCTTC GTGAAGCTGT TTTGAATGGA AAACCAAAC	720
GGAAGTACAT TCAGGCGATT TTGAGAACT GGGCCCATGA AGGAATCAAG AGTGTGGCTC	780
AAATGAGGC CAAGAGAGCA GAAAGAGAAG CAAGCAATCC TCAGTTGACA CAGGTATCTG	840
CAGATTTTAT AAATGCCATG GATCTCTGGA AGGATTAATC CATGCAAGTA GGCTTGAAAT	900
CCGAGTAAGA TTTGCAAGCT GTGTATAATT GTGATAGAAT AAATAGAAAA TAAATTGAAA	960
AAAGAGGTAT GTGAAATGTC ACGTAAACCA TTTATCGCTG GTAAGTGGAA AATGAACAAA	1020
AATCCAGAAG AAGCTAAAGC ATTCGTTGAA GCAGTTGCAT CAAAACCTCC TTCATCAGAT	1080
CTTGTTGAAG CAGGTATCGC TGCTCCAGCT CTTGATTTGA CAACTGTTCT TGCTGTTGCA	1140
AAAGGCTCAA ACCTTAAAGT TGCTGCTCAA AACTGCTACT TTGAAAATGC AGGTGCTTTC	1200
ACTGGTGAAG CTAGCCACCA AGTTTGTAAA GAAATCGGTA CTGACTACGT TGTATCGGT	1260
CACTCAGAAC GCCGTGACTA CTTCCATGAA ACTGATGAAG ATATCAACAA AAAAGCAAAA	1320
GCAATCTTTG CGAACGGTAT GCTTCCAATC ATCTGTTGTG GTGAATCACT TGAACCTTAC	1380
GAAGCTGGTA AAGCTGCTGA ATTCGTAGGT GCTCAAGTAT CTGCTGCATT GGCTGGATTG	1440
ACTGCTGAAC AAGTTGCTGC CTCAGTTATC GCTTATGAGC CAATCTGGGC TATCGGTACT	1500

720

GGTAAATCAG CTTACAAGA CGATGCACAA AAAATGTGTA AAGTTGTTGG TGACGTTGTA 1560  
GCTGCTGACT TTGGTCAAGA AGTCGCAGAC AAAGTTCGTG TTCAATACGG TGTTCTGTT 1620  
AAACCTGAAA ATGTTGCTTC ATACATGGCT TGCCAGACG TTGACGGTGC CTTGTAGGT 1680  
GGTGCCTCAC TTGAAGCTGA AAGCTTCTTG GCTTTGCTTG ACTTTGTAAA ATAATCAGTA 1740  
AGTAGCAAAA GCTAGGTGGA ACAGCATTCA GATGCTGTT ACATTTTTTA TAGGAGAGAA 1800  
AGATTGAAAA CAAAAATTGG ATTAGCAAGT ATCTGTTTAC TAGGCTTGGC AACTAGTCAT 1860  
GTCGCTGCAA ATGAACTGA AGTAGCAAAA ACTTCGCAGG ATACACGAC AGCTTCAAGT 1920  
AGTTCAGAGC AAAATCAGTC TTCTAATAAA ACGCAACGA GCGCAGAACT ACAGACTAAT 1980  
GCTGCTGCCC ACTGGGATGG GGATTATTAT GTAAAGGATG ATGTTCTAA AGCTCAAAGT 2040  
GAATGGATTT TTGACAACTA CTATAAGCCT TGGTTTTATA TTAATTCAGA TGTCGTTAC 2100  
TCGCAGAATG AATGGCATGG AAATTAATAC CTGAAATCAG GTGGATATAT GGCCCAAAAC 2160  
GAGTGGATCT ATGACAGTAA TTACAAGAGT TGGTTTTATC TCAAGTCAGA TGGGGCTTAT 2220  
GCTCATCAAG AATGGCAATT GATTGGAAAT AAGTGGTACT ACTTCAAGAA GTGGGGTTAC 2280  
ATGGCTAAAA GCCAATGGCA AGGAAGTTAT TTCTTGAATG GTCAAGGAGC TATGATGCAA 2340  
AATGAATGGC TCTATGATCC AGCCTATTCT GCTTATTTT ATCTAAAATC CGATGGAAC 2400  
TATGCTAACC AAGAGTGGCA AAAAGTGGGC GGCAAAATGGT ACTATTTCAG GAAGTGGGGC 2460  
TATATGGCTC GGAATGAGTG GCAAGGCAAC TACTATTTGA CTGGAAGTGG TGCCATGGCG 2520  
ACTGACGAAG TGATTATGGA TGGTACTCGC TATATCTTTG CGGCCTCTGG TGAGCTCAAA 2580  
GAAAAAAAAG ATTTGAATGT CGGCTGGGTT CACAGAGATG GTAAGCGCTA TTTCTTTAAT 2640  
AATAGAGAAG AACAAGTGGG AACCGAACAT GCTAAGAAAG TCATTGATAT TAGTGAGCAC 2700  
AATGGTCGTA TCAATGATTG GAAAAAGGTT ATTGATGAGA ACGAAGTGGG TGTTGTCATT 2760  
GTTCTGCTAG GTTATAGCGG TAAAGAAGAC AAGGAATTGG CGCATAACAT TAAGGAGTTA 2820  
AACCCTCTGG GAATTCCTTA TGGTGTCTAT CTCTATACCT ATGCTGAAAA TGAGACCCAT 2880  
GCTGAGAGTG ACGCTAAACA GACCATTGAA CTTATAAAGA AATACAATAT GAACCTGTCT 2940  
TACCCTATCT ATTATGATGT TGAGAATTGG GAATATGTAA ATAAGAGCAA GAGAGCTCCA 3000  
AGTGATACAG GCACTTGGGT TAAAATCATC AACAAGTACA TGGACACGAT GAAGCAGGCG 3060  
GGTTATCAAA ATGTGTATGT CTATAGCTAT CGTAGTTTAT TACAGACGCG TTTAAACAC 3120  
CCAGATATTT TAAAACATGT AAAGTGGGTA GCGGCCTATA CGAATGCTTT AGAATGGGAA 3180  
AACCCTCATT ATTCAGGAAA AAAAGGTTGG CAATATACCT CTTCTGAATA CATGAAAGGA 3240  
ATCCAAGGGC GCGTAGATGT CAGCGTTTGG TATTAAGCGA TGATTTGAAA GAGGGATGTG 3300



721

ATAGTAGCAC CCTCTTTTTC TTTGTTTAT GATAGTTCAT CCTCGAGTAA ATTCAAGTTC 3360  
 TTGCTCGGAA ATGAAGCTTA TATAGTAGAT TGAATATAGA CAAATACCTT GTGATTGGTA 3420  
 AAACATTTTA GAAATTCATT TACCTTTCCT AATCGACTTG GTTTCATCTT ATTTCAATCT 3480  
 ATTATAGTAT TGGGGAATTT CTTCAAACCA CATCAGCTTG GTCAGTTCTA CCTGCGACCT 3540  
 CAAAACCTTG GCTTTGGTCA AGCTGGGTTT AGTTTCCTAG TTTGCTGATG GATTTCCATT 3600  
 GACTATAAGC ATCCAACCCT CTTTTGTCT TCTAAAGAAT TCTTAAATTA TCAGTCTATT 3660  
 GCAACTTTTC TCATATAAGT TCTTTGTCTT GCTATTGGTT TTCCTTAGTA GTATACTAAG 3720  
 GTAGTAATCA TTAAGAAGTG GTTACAAAAA ATAATGAATG AGGTAAAGAA AATGGTAGAA 3780  
 TTGAAAAAAG AAGCAGTAAA AGACGTAACA TCATTGACAA AAGCAGCGCC GG 3832

(2) INFORMATION FOR SEQ ID NO: 93:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10690 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

TGAAAAAATC CTCATGAACC TGGCGCCAAT AGACAAGTGT CTTGTTTCCC TCACCTTCCT 60  
 TATAGGCATG GTCAGCTGAC ACTCGATTGA AGGGTTTAA AGAAACCTTT GTAATTTCTGA 120  
 CAATGCAGAC AGCCTGATTT TGACTATCTA AAATGACATC GAAGGTCCCT ACTTGGGGAA 180  
 GTGGTTCGTC TTCTAGCACA TAGAGGTCAT AGGCTGATGC TGTGCTGTC TTTTCTCCTT 240  
 TAAACACCAA ATCCGCTAAA AGGTCTGGTT CAACTCCAAA AGCCCAGGCA TCGATTTTCAT 300  
 CTCCGATCAA AGGATTGATT TGCTTGATT TATTCACAT TTCTTGCGGT ATCATGGGTG 360  
 CTCCTTTGTA ATTTTACT TTCTCTTTT ATGTGTTTAA GATGATCTGG ATGGTCAATC 420  
 TCTAAATCAA AAATCTCTGG AATAGAACTG TAGTGGATAA TGCACCTGAT ACCCAACTGA 480  
 TTCATTTTTT GTATGAAAGA AGTATTCAGA TAGCCTGCTA CAGCAAAATC AATCTTGTTT 540  
 TTTCTTGCTT TATCCTGCAT ATCTCTTAGC ATATCTAACA TTATTGGACT TTCCATATCA 600  
 TGCCATTGAC TGTTTCTCAT AGTCGCAAAA ACAAGGAAG TCAAATCATT CATTCCAAC 660  
 ACAATCTTTG AAATGCCCGT TTCCAGTATA CTAGATAAGT CAAAATACGC TGACGGTAAT 720  
 TCAATCATCG TTCCGACTTT CCCAGTAAAA CCCTGCTGAC GCAATACTGT AATAGCTTGT 780  
 TTTAATTGGT CGGCATCATT GACAAAAGGA AAGATAACAG ATAGATTGGG GTTGGTTTGA 840

722

TAAACTTCTG TAACGACATG TGCTTCAGCC TGAAATTCAT CCAAACACGC CAGTAAACGC	900
CTAGTTCCTC TATAGCCAAA CAAGGGATGC CCTTCGTCAA AAAACTCTTT AGTCCCCACT	960
AAACAATTGG CTTCTGTATT CGTTAATTCA GTAAAACGAT ACCAAACTTC CTTACCTAAG	1020
TAAAAGGAGC AAATAGTATC AAGATAATCT TTCACAAATT CCTGACAACT TTGTAATAGT	1080
ATATTTTGAT TGAGCTCTCT CAATAAGTAT TCCCCACGAA <del>TCATGCCGAC</del> GTGGTGAAAT	1140
AGTTGAGGAT AAATTTTTTC AAGAAATTTT TCGCCACTAA GGGCAAGTTG ATTTCTCATC	1200
ATTACCTTC CAATTCATGT AAGAAGTCTT GTCCAGTTCT GGAAATCCTA ATAATTCAGA	1260
CTTAACCTTC AAGACTAATG GCGATGCATT TTCTTCTGTA ATCTCTTGAA TATCCATCCA	1320
AATATATCCA AGTGAATCAT TCGCACCATC AGACACAGCT TCCGAAATCG TAACTTGAGG	1380
TGCACTCTCA TTCATTTCOA CATCATACAA GGCTATGACA TGGTGAACCA TAAAATTTTT	1440
TAACTCTTCC CTGACGAAAA CATCGTAGAT TCGAGGATTA GAGTAGCTTC TAACAGTAAA	1500
TCCCCTCTCT TCCATAACTT CTCTAGTCAG CGTTTCCGTC AGTCCTTCAC CAAGTTGCTG	1560
ACTGCCTCCA GGTAGATCAT ACCGATGTTG ATAAGGGCCT CTCGTTTTTT CAATGCAAAG	1620
TAACTTTCCA TTTTCAAAGC AAACACAGTA GACCCCAAAG TGATTTTTGA TTTCCATCCA	1680
ACTCCTCCTA CTTCAAAGAC CAGCCACCAT CTATTGTCAA GATTTGTCCT TGCATGGCGC	1740
TCGCTTTTTCC ACTTGCTAAA AAAAGACTAA GCTCTGCTAT TTCCTCTGGC TCAATCCAGC	1800
GCTTGATTGG GGTTCACATA GCCACCCAGT CAGCCAAACC ACCTGGTTCA AAATCCGCAG	1860
CGGTCATAGC TGTCTTGACT GTCCTGGAG CGATACCCAA GACCTGAATC CCAGCTTCAG	1920
CATAGTCTAG AGCCAACTGC TTGGTGAAGC CAGCCAAGGC ATGCTTGGAT GAAGTATAGG	1980
CGTGACCACC TCCACCTGCT AGGCTAGAAG CAATGGAACA CATATTGATG ATGATTCCCT	2040
TTTTATTTTC CAGCATTTGT GTCAAATAAT ACCGAGTCAA CTCTACTGGA ATAATGTAGT	2100
TGATTTCAAA AATCTCTTGA ATGTCCTGCG CCGTTTGTTC CAACAGTGGT TTGTAATCAT	2160
CCAAAACCTC AGCAGTATTA CAAAAACAT CCACCTGAGG GCACCACTCA AAAATAGGTT	2220
CCAAGTCCAA GGTCAAATCT CTCTGTAAAA AGCGAAAAATC ACCCTCTAAG AGTGGCTTTT	2280
CACCTTGGTC AACTCCATAA ACTTGATAGC CCTTCTCTAA AAAGAGGCGA GCTTGAGCCA	2340
ATCCGATCCC TGAATCACT CCTGTAATGA GTACACGTTT AGTCATGCAC TTCTACCCAA	2400
TCCGTTGCCA AAACATCACA AACTGTCGGG CTCCACATGG AAAAACCTTC TCCTTCGCCA	2460
GAAACGTTGA TTAGGAAATA AGGTGTCATT TCAAGTGCAA GCCCATTTTG CTCGATGGTA	2520
TCAAAGAGTT GGACATAGTT TTCCGCACCT CCCCACCCAG TTCGTACATA TTTTCTCTTA	2580
GCCTTTAACC CAGGCAGGAT CTCTCAAAT GTCATGTTTT TCTCCTTTAA TTCTACATTC	2640

723

TTCATTTAAT TATAGCAAAA AACCGCTTTA TACGGCTTTT TGAATGTGAG TTATTCAAAC	2700
CTGCTACTAC TTACGGCAAA TTATTCCTG CAGCAAGATA AATTCATAC CATTCTTTTC	2760
TTGTTAAGCT AAAGTTTGCC GCTCGGCTAA CTTCTCTCAA GTGCTTAGGA TTTGTTGTAC	2820
CTACGACTGC CTGCATTTT GCTGGATAAC GCAATATCCA AGAAATGGCA ATAGTTGAAG	2880
AGGTTACTCC ATATTTAATA GCTAAACGAT CAAGTACTTG ATTTAAAGCT TGAAATTTCT	2940
CATTTCCAAC AAAATTCCCT TTAAATACC CGAATTGTAA GACAGACCAT GCTTGAATGA	3000
CCACATCGTG TAATTGGCAA TATTCAAAA TGCTGCCATC TCGCATAGCT GCTTGAATAT	3060
CTTCCATATT AACATGAAAA GCTGATTCAA ATCCTGGAGT AAAAGCCGCA CTCAATTGTA	3120
GCTGATTAAC AGCTAACGGC TGCTTGACAT CTTTTTAAAG CAACTCCATC ATCATAGGAT	3180
TTTGATTAGA AACTCCAAAA TCTCGAACTT TACCTTGTTT ATAAAGGAGA TTAAAGGCTT	3240
CTGCTACTTG GTCAGATTCC ATCAAAGCAT CTGGTCGATG AAGGAGCAAG CTATCTAGAT	3300
GATCAATCTT CAATCTTGC AAAATACCGT CACTGATTT TATAATATAG TCCTTAGAAA	3360
AATCAAAATA GGTAAATTCT TCAATGCGAA TGCCACATTG GACTGAATC CACATCTTTT	3420
CTCTTAAATC TGGACGATT TTTAGGACAA GACCTAACAG TTCTTCACAA CGACCACGAC	3480
CATAAATATC AGCCAAGTCG AAGGCATTGA TTCCAACAGA AAGTGCTGTT TCTACAAGCT	3540
CTTCAACTTC TTTTACAGAT TTATCTTTTA TTCTCATCAT TCCGAGAACA ATTTCTGATA	3600
ATTCTTTGTC ATCTTGACCA AGAGTTATGT ATCTCATCAA ATTTTCTCC TTTAATTTCT	3660
AACATTCTTC CCTTCATTAT AACAAAAAC CGCTTTGCAA CGACTTTTG ACTATACTTC	3720
ACTCCATTTT ATCTTCTTAA ACCCACGGAA CAAGACAAAG ATTCCAATAA AGAGGACAGC	3780
TAAAGGAATA ACTTTTGTA GGAACATT TGAAATTTCC ATCCACTCAT AATAACGGAG	3840
CAGAGAACCC ACCACAAGAT GGGCAATAAT CATACTGACA AATGGACGAA AGACCGCTTC	3900
TTTCCAATTC CAAATACCGA TAACTAGCGA AATCGTAAAG ACAGACAAAC TATCCCAGGG	3960
AGCCGGAATA TAAAAGGCTC CTTCTTGAT GAAGCTTGCC ATTCCTACAT ATCCTAAAAC	4020
AACTAGAAGA ACTATAGTCC CAACAACAAT GTAAGTGCCA ATTTTCATTT TAGGAGAATC	4080
TTGGACTAAA CTTCTTCGTA AAATTGTGGC CACAAGTCCA AATCCAATCA GAAAAATAAG	4140
AAGTTGCCCT AAAAATGTGA GCAAATTGAC TGTTAAGAGA GGACCTTTAG AAAATCACT	4200
TAGTAGTTGA TAATAACGTA ATACCGCCAG GACAAGAATT GCGTCAAAA GGGACTCTTT	4260
GATAGAAGTG CGAGGTGCTC CTTGAGAAT CTCTTTCATT ATTTTTTTAG GATTCTTACC	4320
TAGATAATCC TCTGCACTCA TGCCATCTCG TTCTGCTTCT GAGAAATCTA GCATCATCAA	4380

724

ATAGATCTGC TCTCTGAGAT AGTCTTCATC ATAGAGAAAT CCAGCAAGAT TAAAACTTTC	4440
CCACAACCTCC TCAAAATACT TTTGATTCTC CTCAGAAAAC TCATGTAGCA AAGCGCTTGT	4500
TTCTTCGTAA TACTTCATTT TCTTCATGGT TTAACCCCA TTCTTAATCC CTTCTACTTT	4560
TTGACTCAAA TCGTCCCAT TTTGCCAAA GACTGAGACA CGCTCTTCTC CTTCTTTCAT	4620
TAATGAAAA TACTTCCGAT CTGGACCATC TGGCGACGGG CGCATGTGCG CTCTTATCCA	4680
TTGATTTTT TCTAACTTTT GCAACAAAGG ATAAATAGTT CCTGGAACGA TAGTATCAAA	4740
TCCAGCCTCT CGCAAAGTCT GAACCAACTC ATAACCATAC CGCTCTTTT GACCAATCAT	4800
ATCCAAGACA CAACCTTCAA GAACACCTTT TAATAGCTGA GTTCTTTCA TCACTTCTCC	4860
CTTCTAATCT ATTTTGTAAT ACCTACTAGT GACTTCACCT ATAGTATATC ACTTCTACAC	4920
TAGTTTGTA AGCATAATAG TTAATACTCT TCGAAAATCT CTTCAAACCA CGTCAGCGTC	4980
GCCCTACCGT ATGTATGGTT ACTGACTTCG TCAGTTTCAT CTACAACCTC AAAAACATGT	5040
TTTGAGCTGA CTTGCTCAGT TTCATCTACA ACCTCAAAC AGTGTTTGA GCTGACTTCG	5100
TCAGTTTCAT CTACAACCTC AAAACAGTGT TTTGAGCTGA CTTGCTCAGT TTCATCTACA	5160
ACCTCAAAAA CATGTTTGA GCTGACTTCG TCAGTTTCGT CTACAACCTC AAAACAGTGT	5220
TTTGAGCAAC CTGCGGCTAG CTTCTTAGTT TGCTCTTTGA TTTTCATTGA GTATAAATAA	5280
AAAAACAGAA CTAGCCTGAA CTAGTCTGT CTACTTTTAC CCAATCACAC TTCCATTGG	5340
TACAGCTGGA TCAACTGTGA GAAGGGTTAA TTTGCCATCA TGTTTCACTG AGAGAATCAT	5400
ACCCTGGCTG ACATATTTTT TCATCATTTT ACGTGGTTTG AGGTTAGCAA CGATTTGAAC	5460
TTTCTTGCCG ACCAATCTT GTTCATTTGG ATAGTATTTT GCAATTCCTG AAAGAATCTG	5520
ACGATCTTCT CCATCACCAG CATCCAAGCG GAATTGAAGC AACTTATCTG AACCTTCTAC	5580
TTTAGACACT TCTTTGACTT CTGCGACACG GATTTCAACC TTGTCAAAGT CTTCAAACCT	5640
GATTTTCATCC TTGTTTAGTT TGAGCTCAAC TTCGTCCGGA TTCCATTCTT TTTCGACTGC	5700
TGGTTTATTG CTTTCCATTT GTTCCTTGAT ATAGGCGATT TCTTCTTCCA TATTTAGACG	5760
TGGAAAGATA GGTGTTCTT TGGCAACTAC AGTCACATCT GCTGGGAAGT CAGCCAACT	5820
CAAGTTTCA AACTAGAAA CTTCTTCCAA ACCAAGTTGA GTCAAACTG CACGACTAGT	5880
TTCCATCATA AATGGTTCOA TCAAGTGAGC AACTACACGA ATGCTGGCTG CCAAGTGGCT	5940
CATGACACTT GCCAATTGGT CACGAAGAGC TTCATCCTTG GCCAAGACCC ATGGTGGCTG	6000
CTCATCGATG TATTTATTGG TACGAGAGAT CAGAGTCCAG ACTGCTTCAA GCGCACGTGG	6060
ATAGTCAACT GCTTCCATGT GTGTATGGAA GTCTGCGATT GATTGTWCTG CAACCTCAGC	6120
AAGAACATGA TCATATTCAG TCACACCTTC TACATAGGCA GGGATTTGTC CATCAAAGTA	6180

725

CTTATTAATC ATGGAACCG TACGGTTAAG GAGGTTCCCA AGGTCATTAG CCAATTCATA	6240
GTTGATACGG CCGACATAGT CTTCAAGGAGT AAAGGTTCCG TCTGAACCAA CTGGAAGGTT	6300
ACGCATGAGG TAGTAACGAA GTGGATCTAG TCCATAACGC TCTACCAACA TTTCAGGGTA	6360
AACGACATTC CCTTTTGACT TAGACATTTT TCCGTCTTTC ATGACAAACC AACCATGGGC	6420
AATCAAACGA TCAGGTAATT TAACATCCAA CATCATAAGA AGGATTGGCC AGTAGATAGA	6480
GTGGAAGCGA AGGATATCTT TTCCTACCAT ATGGAAGACT GTTCCATTCC AGAACTTGTC	6540
AAAGTTACCA TGTTCTGCTT GAGCGTAGCC AAGAGCTGTC GCATAGTTAA GAAGGGCATC	6600
AATCCAAACG TAGACAACGT GTTTTGGATT TGATGGGACA GGCACCTCCC ATGTAAAGGT	6660
TGTACGAGAT ACCGCCAAAT CTTCCAAGCC TGGCTCGATG AAGTTGCGTA GCATTTTCATT	6720
AAGGCGACCA TCTGGCGTGA TAAATTCAGG ATGAGCTTTG AAAAATTCGA CCAAACGGTC	6780
TTGGTATTTG CTAAGGCGAA GGAAGTATGA TTCTTCAGAA ACCCATTCAA CCTCATGACC	6840
TGATGGAGCA ATACCACCAG TCACATTTCC AGCTTCATCA CGGAAAACCT CTGCCAGCTG	6900
GCTTTCTGTA AAGAATTCCT CGTCTGATAC TGAATACCAA CCAGAGTATT CACCCAAGTA	6960
GATATCATCT TGAGCAAGTA AGCGTTCAAA GACTTGTGCG ACAACTTTTT CATGGTAGTC	7020
ATCAGTTGTA CGGATAAATT TATCGTATGA GATATCTAGT AATGCCAGA GTTCTTTAAC	7080
TCCAACCGCC ATTCCATCAA CATAGGCTTG AGGTGTAATA CCAGCTTCTT CCGCTTTCTG	7140
CTGGATTTTC TGACCATGTT CATCAAGACC TGTGAGATAA AATACATCGT AGCCCATCAG	7200
GCCTTTGTAA CGTGCTAGGA CATCACATGC GATAGTTGTG TAGGCAGAAC CGATATGAAG	7260
TTTCCCAGAT GGATAGTAAA TCGGCGTTGT AATATAAAAA TTTTTTTCAG ACATAATTTT	7320
TCCTTTCCAG GCAAATGAAA CCTGTTTTTC TAACACTTCA TTATATCACA TTTTAAATGA	7380
ATTTCAATAG GGAAATCCAT AAAAAACAA GATAGACGAG TGTCCATCTT GTTGATCTCA	7440
TTCATAACGA AGGGCTTCAA TTGGATCAAG TTTTCATGCC TTGTTGGCTG GCAAGACTCC	7500
AAAAATCATA CCAACACTAG CCGAAACTGC AAGACTAAAT AGGGCGACTG GGATTGATAC	7560
TCCAACCTCT ATACCTTCTA TTAACCTTG CAGTAACAAA CCTGCTAAGG CAGTTAAACC	7620
ACTTGCAATT GTCAAGCCAA TTAAGCCACC TAACAAGGTC AAAATCATGG ATTCAATCAA	7680
AAACTGAATT AAAATATTGG CACGTGTTGC ACCCAAAGCC TTACGAAGAC CAATCTCACC	7740
AGTGCGCTCT GTCACCGAAA CCAGCATGAT GTTCATGACA CCAGTTCCTC CAACAAAGAG	7800
AGAAATCCCT GCGATGGAAC TAATAATCGT CGTCATAAAA CTAAACGATT GTTGAATTTT	7860
TGCAAATACA ACGGACTCAT CTGCCACCTG GTATTCTCCC TGTGTGAAGC CTGCAAGCTC	7920

726

TGTCATTTTT	CGTGCCAGTT	CTGGACCCAG	AGTTGGGGTT	AAACTGGTAT	CATTCACTCG	7980
AAAGACAATA	TTAGCTATTT	CATCTACATT	AAAATTCGCA	GCAAGGGAGA	TATTGGTAGT	8040
AATAGGCAAG	CCACCAAACC	CATATATTTT	TGATCTTTTA	GCCTCCGGAC	TAGTATAAAC	8100
CCCAATGACC	CGGTAACATA	ATCCATTGAC	TTCTACAACC	TTGTAAATAG	CCTCTTGAGG	8160
AGATTCAAAT	AAACTAATGG	ACAATTCCTC	ATCTAGCAAA	ATGACACTTG	CAAACCTCTT	8220
GAAATCTTGC	TCTCTCAGAC	TACGACCTGC	AATAATTTCA	TTCTTAACAG	CGTCCATGTA	8280
AGTTCTGTTT	CCACCTGTCA	AATTAGCATT	CTCAACCTTT	TTATCTTGAT	AGGTCAAGAT	8340
GGCATTCGTT	GAATTGGTTA	CATAGTAACT	ATCCACTCCC	TTCAGTTTAG	CTGCCTCTTG	8400
GACCCAGGAT	TCTTGCGGTT	TTGGCGGTTT	AACAGGAACT	TCCTCTTCCT	TTCCAGAAAC	8460
CGTAAAAGCT	GATTGTTTCT	GAGTAAAAGA	CCCGTCTTTA	CTTTTTTTAG	GAGAGAAAAA	8520
GACGCTAATA	TTTTTCTGAG	ATTTAGTCAT	ATCTTTATTG	ACTTGACGAG	ATAGGGAATC	8580
ACCCAAAGCC	ATAATCACAA	CAACTGATGA	AACACCGATA	ATAATCCCAA	TCATAGTAAG	8640
CAAAGAACGC	ATCTTGTGAG	CCATGATAGA	TGAAAAGGCA	AATTCAGAT	TCTGCATCTT	8700
AGTTTTCTCT	CTTTCCTAAC	TGAGCACTGT	CAGACGAAAT	GACCCCATCC	CGAATGACAA	8760
TCTGACGTTT	GGCATAGGCA	GCAATCTCAG	GCTCATGCGT	TACCATGATA	ATGGTTTTTC	8820
CTTCTTTATT	CAAATCAACC	AATAATTGCA	TAATTTGGTT	ACCTGTTTTG	GTATCCAAGG	8880
CTCCTGTCGG	TTCATCCGCT	AGGATAATAG	AAGGATTGTT	TACCAAGGCA	CGCGCAATGG	8940
CTACACGTTG	CTTTTGACCA	CCAGATAATT	CTGAAGGTAA	ATGGTGACTA	CGTTCTGTCA	9000
ATTCAACCTT	GTCTAAATAT	TCCTCAGCCA	ACTTGCGACG	TTTTGAAGAC	GAAACTCCTG	9060
CGTAAATCAA	GGGCAATTCT	ACATTTTGCA	GAGCATTGAG	CTTCGATAGA	AGAAAGAACT	9120
GCTGAAAGAC	AAAACCGATT	TGTTGGTTAC	GGACCTTAGC	TAGTTGTTTT	TCACCAAGCC	9180
CAGCCACTTC	TTGACCTTCA	AGATAATATT	CTCCACTGGT	TGGTGTATCC	AACATGCCAA	9240
TCGTATTCAT	CAGAGTGGAC	TTACCAGACC	CAGATGGTCC	CATGATGGCT	ACAAATTCAC	9300
CCTCATTCAC	TTCTAGATTG	ATATTTTGA	GAACCTGCAG	TTCTTGGTCA	CCATTACGGT	9360
AACTTCTGAA	GATATTTTTT	AGACTAATTA	GTTGCTTCAT	CAGCCTTCAC	CTCTTTTCCT	9420
TCTTCCAAGG	AAGATGTTGG	ATTACTGATG	ACCTTAGCAC	CGTTCGTTAA	ACCAGAAGTG	9480
ATTTCTTGAT	TTCTGCGTC	AGCATTTCCC	AATGAAACCT	CAACTTTTTT	AGCCTTTTGT	9540
TGTTTCATCA	CAATCCAGAC	ATAATTTTAA	CTATCATCCA	TTACTAGACT	GCTAACAGGA	9600
ACAAGAATAG	CCTTAGTTTT	GCTTTTAAAC	TCAATGTTGA	CAGAAAAACC	TTGTTTCAAA	9660
TCACCAACCT	CGCCTGTCAC	ATCAATAGTA	TAAGGGTATT	TAGAACCTGT	ATTATTCCCG	9720

727

GCTGCTGGAC TAGCTGCTTC ACCATTGTTT TTAGGATAGT CAGAAATATA GCTTAATTTT	9780
CCAGTCCATT TTTTATCAGG ATACACTTTA GAAGTAAAGC TTACTTCTTG ACCTACAGAA	9840
AGGTTGGCTA GATTGTACTC AGACAATTCT CCCTTGACTT GTAAATTTTC ATTGCTGACA	9900
ATATGAACCA TAACTTGACT CGCCCTGTT GGAGATTAG AACATTGCT ATTGACTTCG	9960
ACCACAGTTC CCTCTAGGGT ACTGAGAACA GTTGTGTCAT CCAATTGACT TTGAGCCTTG	10020
CTTAATTGCG CCGCAGCATC TGCACGCGCA TCACGGGCAT CACCCAATTG AGCGTCAATA	10080
GAAGCAACAG AATTTCAGC CACTGGAGTT GGGCTTTGCA CCGTTGCATC TTCTCCTCCT	10140
ACTGGCGCTG GTAAGTGTGG AGCCGGAGCT GAAGCGGCTT CATTTCGTGC TTGATTGAGT	10200
TCATTGATAT GACGATCTGC CCTAGCTACT GCTCGACTAG CTGAATCATA GGCCGCTGC	10260
GCTTCTGAAC TACTGTACTT GACTAAAGCC TGCCCTTCGC TGACCTTATC GCCCAGAGAA	10320
ACAAGGATTT CATCTAAATC ACCCTTACTA GCATCAAAAT AACATATTG TTCATTTTTT	10380
GCTGTTACTG TCCCTGACAA TAAACAGAG GAGGCCACGC TTCCTTCCTT GGCAACAACA	10440
AGATGAGTAG GCTCATCTTT TAGAGCAGTC TGAGAAGGTT GTCTAAAGAG TAAATCCCC	10500
CCAGCACCCA ATACAACTAC ACTCGCAGCA CCGATTGCTG CATACAGTTG CCACTTTTTA	10560
GCTTTACCAT TCTTTTCTT CATAATGAAA CTCCTTTTCT TTTTACAAT ACTTTGCTAT	10620
TATACCAAAT TTCCCTCCAG CAAACAATAC AGTTCAGGAT TAAACAATCG TTCGGAATTT	10680
TGCTTTTCGG	10690

## (2) INFORMATION FOR SEQ ID NO: 94:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

GAGAAAGCGC CCACGTTTCC CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCCA	60
GGGTCGGAAC AGGAGAGCGC AACGAGGGAG CTTCACAGG GGAACGCCT GGTATCTTTA	120
TAGTCCTGTC GGGTTTCGCC ACCTCTGACT TGAGCGTCGA TTTTGTGAT GCTCGTCAGG	180
GGGGCGGAGC CTATGAAAAA ACGCCAGCAA CGCGGCCTTT TTACGGTTCC TGGCCTTTTG	240
CTGGCCTTTT GCTCACATGT TCTTCTGTC GTTATCCCCT GATTCTGTGG ATAACCGTAT	300
TACCGCCTTT GAGTGAGCTG ATACCGCTCG CCGCAGCGCA ACGACCGAGC GCAGCGAGTC	360

728

AGTGAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC	420
GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA	480
CGCAATTAAT GTGAGTTAGC TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC	540
GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA	600
CaTGATTACG AATTCGAGCT CGGTACCCGG AAAATCCAGA AAATGCTTGA AAAAAATCCT	660
AGAAGATGGT ATAATACTAA ATTGTAAGGG TTATCACATA TAACTCAAAA AAAGAAAGAA	720
CAAAAGGAGA GTCAAACTAT GGCTTCTAAA GATTTCACG TAGTGGCAGA AACAGGTATT	780
CACGCACGTC CAGCAACATT GTTGGTACAA ACTGCTAGCA AATTTGCTTC AGATATCACT	840
CTTGAGTACA AAGGTAAATC AGTTAACCTT AAATCAATTA TGGGTGTTAT GAGTCTTGGT	900
GTTGGCCAAG GTGCTGACGT AACTATCTCA GCTGAAGGTG CAGATGCAGA TGACGCTATC	960
GCTGCAATCT CAGAAACAAT GAAAAAGAA GGATTGGCAT AAGGGAAATG ACAGAAATGC	1020
TTAAAGGAAT CGCAGCATCT GACGGTGTG CAGTTGCAAA AGCATATCTA CTCGTTCCAGC	1080
CGGATTTGTC ATTTGAGACT ATTACAGTCG AAGATACAAA CGCAGAAGAA GCTCGCCTTG	1140
ATGCCGCTCT ACAGGCATCA CAAGACGAGC TTTCTGTTAT TCGCGAGAAA GCAGTAGGTA	1200
CGCTCGGTGA AGAAGCAGCT CAAGTTTTTG ATGCTCACTT AATGGTTCTT GCTGACCCAG	1260
AAATGATCAG CCAAATCAAG GAACTATCC GTGCGAAGAA AGTGAATGCA GAAGCAGGTC	1320
TGAAAGAAGT TACAGATATG TTTATCACTA TCTTTGAAGG CATGGAAGAC AACCCTATACA	1380
TGCAAGAACG CGCAGCGGAT WTCGCGACG TGACAAAACG TGTATTGGCA AACCTTCTTG	1440
GTA AAAAATT GCCAAACCA GCTTCTATCA ATGAAGAAGT GATTGTGATT GCGCATGACT	1500
TGACTCCTTC AGATACAGCT CAATTGGACA AAAACTTTGT AAAAGCTTTT GTAACCAACA	1560
TTGGTGGACG TACAAGCCAC TCAGCTATCA TGGCACGTAC ACTTGAAATT GCTGCTGTAT	1620
TAGGTACAAA TAACATCACT GAAATCGTTA AAGACGGTGA CATCCTTGCT GTTAACGGGA	1680
TCACTGGAGA AGTGATTATC AACCCAACAG ATGAACAAGC GGCAGAATTT AAAGCAGCTG	1740
GTGAAGCCTA TCGGAAACAA AAAGCTGAAT GGGCACTTTT GAAAGATGCT CAAACAGTGA	1800
CTGCTGACGG TAAACACTTC GAGTTGGCTG CTAATATCGG TACTCCAAAA GACGTTGAAG	1860
GTGTTAACAA CAACGGTGCA GAAGCTGTTG GACTTTACCG TACAGAGTTC TTGTACATGG	1920
ATTCTCAAGA CTTCCCAACT GAAGATGAGC AGTATGAAGC ATACAAGGCT GTTCTTGAAG	1980
GAATGAACGG TAAACCTGTT GTCGTTCTGA CAATGGATAT CGGTGGAGAT AAGGAACTTC	2040
CTTACTTCGA TATGCCTCAC GAAATGAACC CATTCCTTGG ATTCCGTGCT CTTCGTATCT	2100
CTATCTCTGA GACTGGAGAT GCTATGTTCC GCACACAAAT CCGTGCTCTT CTTGCTGCGT	2160



CTGTTACACGG TCAATTGCGT ATCATGTTCC CAATGGTTGC GCTCTTGAAA GAATTCCTGTG	2220
CAGCGAAAGC AGTCTTTGAT GAAGAAAAAG CAAACCTTCT TGCTGAAGGT GTTGCAGTTG	2280
CGGATAACAT CCAAGTTGGT ATCATGATCG AGATTCCTGC AGCGGCTATG CTTGCAGACC	2340
AATTTGCTAA AGAAGTTGAC TTCTTCTCAA TTGGTACAAA CGACTTGATC CAATATACAA	2400
TGGCAGCAGA CCGTATGAAC GAACAAGTTT CATACTTTA CCAACCATAC AACCCATCAA	2460
TCCTACGCTT GATTAACAAT GTGATCAAAG CAGCTCACGC TGAAGGTAAA TGGGCTGGTA	2520
TGTGTGGTGA GATGGCTGGT GACCAACAAG CTGTTCCACT TCTTGTGCGA ATGGGCTTGG	2580
ATGAGTTCTC TATGTCAGCA ACATCTGTAC TTCGTACACG CAGCTTGATG AAGAACTCG	2640
ACACAGCTAA GATGGAAGAG TACGCAAACC GTGCCCTTAC AGAATGCTCA ACAATGGAAG	2700
AAGTTCTTGA ACTTCAAAAA GAATACGTTA ATTTTGATTA ATCGAAAAGT CCCTGCAACT	2760
CAGTTACAGG GATTTTTTTG ATATTTTAAA AAGAATTTT CAGAAAATCT TTCTTATAGA	2820
AAGTCCAACC TTGAAAAAGT AGTGGTCAGA ACAAAAAATA CTTAAATGGT TCATAAAATT	2880
CTTGACAAGT TGGATATTTA GGAGTAACT ATTAACCAGT TAAGTAATAG AGAGGAGTTT	2940
CTGCAATTTA GAAATGAATT GCAACTAGAA ATATCAAATA GAAAGAGAGT TTCGATGAAA	3000
ATTAATAAGA AATACCTTGT TGGTTCTGCG GCACCTTGAT TTTAAGTGT TGTCTTACG	3060
AGTTGGGACT GTATCAAGCT AGAACGGTTA AGGAAAATAA TCGTGTTC TATATAGATG	3120
GAAAACAAGC GACGCAAAAA ACGGAGAATT TGACTCCTGA TGAGGTAGC AAGCGTGAAG	3180
GAATCAATGC TGAGCAAATC GTCATCAAGA TAACAGACCA AGGCTATGTC ACTTCACATG	3240
GCGACCACTA TCATTATTAC AATGGTAAGG TTCCTTATGA CGCTATCATC AGTGAAGAAT	3300
TACTCATGAA AGATCCAAAC TATAAGCTAA AAGATGAGGA TATTGTTAAT GAGGTCAAGG	3360
GTGGATATGT TATCAAGGTA GATGGAAAAT ACTATGTTA CCTTAAGGAT GCTGCCACG	3420
CGGATAACGT CCGTACAAAA GAGGAAATCA ATCGACAAAA ACAAGAGCAT AGTCAACATC	3480
GTGAAGGTGG AACTCCAAGA AACGATGGTG CTGTTGCCTT GGCACGTTG CAAGGACGCT	3540
ATACTACAGA TGATGGTTAT ATCTTTAATG CTTCTGATAT CATAGAGGAT ACTGGTGATG	3600
CTTATATCGT TCCTCATGGA GATCATTACC ATTACATTCC TAAGAATGAG TTATCAGCTA	3660
GCGAGTTGGC TGCTGCAGAA GCCTTCCTAT CTGGTCGAGG AAATCTGTCA AATTCAAGAA	3720
CCTATCGCCG ACAAATAGC GATAACACTT CAAGAACAAA CTGGGTACCT TCTGTAAGCA	3780
ATCCAGGAAC TACAAATACT AACACAAGCA ACAACAGCAA CACTAACAGT CAAGCAAGTC	3840
AAAGTAATGA CATTGATAGT CTCTTGAAAC AGCTCTACAA ACTGCCTTTG AGTCAACGAC	3900

730

ATGTAGAATC TGATGGCCTT GTCTTTGATC CAGCACAAAT CACAAGTCGA ACAGCTAGAG	3960
GTGTTGCAGT GCCACACGGA GATCATTACC ACTTCATCCC TTA CTCTCAA ATGCTCTGAAT	4020
TGGAAGAACG AATCGCTCGT ATTATTCCCC TTCGTTATCG TTCAAACCAT TGGGTACCAG	4080
ATTCAAGGCC AGAACAAACCA AGTCCACAAC CGACTCCGGA ACCTAGTCCA GGCCCGCAAC	4140
CTGCACCAAA TCTTAAAATA GACTCAAATT CTCTTTGGT TAGTCAGCTG GTACGAAAAG	4200
TTGGGGAAGG ATATGTATTC GAAGAAAAGG GCATCTCTCG TTATGTCTTT GCGAAAGATT	4260
TACCATCTGA AACTGTTAAA AATCTTGAAA GCAAGTTATC AAAACAAGAG AGTGTTCAC	4320
ACACTTTAAC TGCTAAAAA GAAAATGTTG CTCCTCGTGA CCAAGAATTT TATGATAAAG	4380
CATATAATCT GTTAACTGAG GCTCATAAAG CCTTGTTTGA AAATAAGGGT CGTAATTCTG	4440
ATTTCCAAGC CTTAGACAAA TTATTAGAAC GCTTGAATGA TGAATCGACT AATAAAGAAA	4500
AATTGGTAGA TGATTTATTG GCATTCTAG CACCAATTAC CCATCCAGAG CGACTTGGCA	4560
AACCAAATTC TCAAATTGAG TATACTGAAG ACGAAGTTCG TATTGCTCAA TTAGCTGATA	4620
AGTATACAAC GTCAGATGGT TACATTTTGG ATGAACATGA TATAATCAGT GATGAAGGAG	4680
ATGCATATGT AACGCCTCAT ATGGGCCATA GTCACTGGAT TGGAAAAGAT AGCCTTTCTG	4740
ATAAGGAAAA AGTTGCAGCT CAAGCCTATA CTAAAGAAAA AGGTATCCTA CCTCCATCTC	4800
CAGACGCAGA TGTTAAAGCA AATCCAACTG GAGATAGTGC AGCAGCTATT TACAATCGTG	4860
TGAAAGGGGA AAAACGAATT CCACTCGTTC GACTTCCATA TATGGTTGAG CATAAGTTG	4920
AGGTTAAAAA CGGTAATTTG ATTATTCCTC ATAAGGATCA TTACCATAAT ATTAAATTTG	4980
CTTGTTTGA TGATCACACA TACAAAGCTC CAAATGGCTA TACCTTGAA GATTGTGTTG	5040
CGACGATTAA GTACTACGTA GAACACCCTG ACGAACGTCC ACATTCTAAT GATGGATGGG	5100
GCAATGCCAG TGAGCATGTG TTAGGCAAGA AAGACCACAG TGAAGATCCA AATAAGAACT	5160
TCAAAGCGGA TGAAGAGCCA GTAGAGGAAA CACCTGCTGA GCCAGAAGTC CCTCAAGTAG	5220
AGACTGAAAA AGTAGAAGCC CAACTCAAAG AAGCAGAAGT TTTGCTTGCG AAAGTAACGG	5280
ATTCTAGTCT GAAAGCCAAT GCAACAGAAA CTCTAGCTGG TTTACGAAAT AATTTGACTC	5340
TTCAAATTAT GGATAACAAT AGTATCATGG CAGAAGCAGA AAAATTACTT GCGTTGTAA	5400
AAGGAAGTAA TCCTTCATCT GTAAGTAAGG AAAAAATAAA CTAATGAAAA ATGAAAGTCT	5460
CGATAAAGAG GCTTTCATTT TTATTATGTA TATATGTAAA ATTCTTGACA AGCAATATTA	5520
AAAAGAGTAA ACTATTAACT AGTTAATTAA CCGGTTTATT ACTTTATAGT GAATCAAATA	5580
TACTTAAGAA AAGAGGAAAG AATGAAAATT AATAAAAAAT ATCTAGCAGG TTCAGTGGCA	5640
GTCTTGCCC TAAGTGTGTTG TTCCTATGAA CTTGGTCGTC ACCAAGCTGG TCAGGTAAAG	5700

731

AAAGAGTCTA ATCGAGTTKc TTATATAGAT GGTGATCAGG CTGGTCAAAA GGCAGAAAAC	5760
TTGACACCAG ATGAAGTCAG TAAGAGGGAG GGGATCAACG CCGAACAAAT CGTCATCAAG	5820
ATTACGGATC AAGGTTATGT GACCTCTCAT GGAGACCATT ATCATTACTA TAATGGCAAG	5880
GTCCCTTATG ATGCCATCAT CAGTGAAGAG CTCCTCATGA AAGATCCGAA TTATCAGTTG	5940
AAGGATTCAG ACATTGTCAA TGAAATCAAG GGTGGTTATG TTATCAAGGT AGATGGAAAA	6000
TACTATGTTT ACCTTAAGGA TGCAGCTCAT GCGGATAATA TTCGGACAAA AGAAGAGATT	6060
AAACGTCAGA AGCAGGAACA CAGTCATAAT CACGGGGGTG GTTCTAACGA TCAAGCAGTA	6120
GTTGCAGCCA GAGCCCAAGG ACGCTATACA ACGGATGATG GTTATATCTT CAATGCATCT	6180
GATATCATTG AGGACACGGG TGATGCTTAT ATCGTTCCTC ACGGCGACCA TTACCATTAC	6240
ATTCCTAAGA ATGAGTTATC AGCTAGCGAG TTAGCTGCTG CAGAAGCCTA TTGGAATGGG	6300
AAGCAGGGAT CTCGTCTTC TTCAAGTTCT AGTTATAATG CAAATCCAGC TCAACCAAGA	6360
TTGTCAGAGA ACCACAATCT GACTGTCACT CCAACTTATC ATCAAAATCA AGGGGAAAAC	6420
ATTTCAAGCC TTTTACGTGA ATTGTATGCT AAACCCCTTAT CAGAACGCCA TGTGGAATCT	6480
GATGGCCTTA TTTTCGACCC AGCGCAAATC ACAAGTCGAA CCGCCAGAGG TGTAGCTGTC	6540
CCTCATGGTA ACCATTACCA CTTTATCCCT TATGAACAAA TGTCTGAATT GGAAAAACGA	6600
ATTGCTCGTA TTATTCCCTC TCGTTATCGT TCAAACCATT GGGTACCAGA TTCAAGACCA	6660
GAACAACCAA GTCCACAATC GACTCCGGAA CCTAGTCCAA GTCCGCAACC TGCACCAAT	6720
CCTCAACCAG CTCCAAGCAA TCCAATTGAT GAGAAATTGG TCAAAGAACC TGTTCGAAAA	6780
GTAGGCGATG GTTATGCTT TGAGGAGAAT GGAGTTTCTC GTTATATCCC AGCCAAGGAT	6840
CTTTCAGCAG AAACAGCAGC AGGCATTGAT AGCAAACCTGG CCAAGCAGGA AAGTTTATCT	6900
CATAAGCTAG GAGCTAAGAA AACTGACCTC CCATCTAGTG ATCGAGAATT TTACAATAAG	6960
GCTTATGACT TACTAGCAAG AATTCACCAA GATTTACTTG ATAATAAAGG TCGACAAGTT	7020
GATTTTGAGG CTTTGGATAA CCTGTTGGAA CGACTCAAGG ATGTCyCAAG TGATAAAGTC	7080
AAGTTAGTGG ATGATATTCT TGCCTTCTTA GCTCCGATTC GTCATCCAGA ACGTTTAGGA	7140
AAACCAAATG CGCAAATTAC CTACACTGAT GATGAGATTC AAGTAGCCAA GTTGGCAGGC	7200
AAGTACACAA CAGAAGACGG TTATATCTTT GATCCTCGTG ATATAACCAG TGATGAGGGG	7260
GATGCCTATG TAACTCCACA TATGACCCAT AGCCACTGGA TTAATAAAGA TAGTTTGTCT	7320
GAAGCTGAGA GAGCGGCAGC CCAGGCTTAT GCTAAAGAGA AAGGTTTGAC CCCTCCTTCG	7380
ACAGACCATC AGGATTCAGG AAATACTGAG GCAAAGGAG CAGAAGCTAT CTACAACCCG	7440

732

GTGAAAGCAG CTAAGAAGGT GCCACTTGAT CGTATGCCTT ACAATCTTCA ATATACTGTA	7500
GAAGTCAAAA ACGGTAGTTT AATCATACCT CATTATGACC ATTACCATAA CATCAAATTT	7560
GAGTGGTTTG ACGAAGGCCT TTATGAGGCA CCTAAGGGGT ATACTCTTGA GGATCTTTTG	7620
GCGACTGTCA AGTACTATGT CGAACATCCA AACGAACGTC CGCATTGAGA TAATGGTTTT	7680
GGTAACGCTA GCGACCATGT TCGTAAAAAT AAGGTAGACC AAGACAGTAA ACCTGATGAA	7740
GATAAGGAAC ATGATGAAGT AAGTGAGCCA ACTCACCTG AATCTGATGA AAAAGAGAAT	7800
CACGCTGGTT TAAATCCTTC AGCAGATAAT CTTTATAAAC CAAGCACTGA TACGGAAGAG	7860
ACAGAGGAAG AAGCTGAAGA TACCACAGAT GAGGCTGAAA TTCCTCAAGT AGAGAATTCT	7920
GTTATTAACG CTAAGATAGC AGATGCGGAG GCCTTGCTAG AAAAAGTAAC AGATCCTAGT	7980
ATTAGACAAA ATGCTATGGA GACATTGACT GGTCTAAAAA GTAGTCTTCT TCTCGGAACG	8040
AAAGATAATA ACACTATTTC AGCAGAAGTA GATAGTCTCT TGGCTTTGTT AAAAGAAAGT	8100
CAACCGGCTC CTATACAGTA GTAAATGAA TGGAGCATAT TTTATGGAGA AGTAACCTTT	8160
CGTGTTACTT CTCTTTTTTA GAAAAACGTA ACAGA	8195

(2) INFORMATION FOR SEQ ID NO: 95:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2004 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

TTTACTAAAA GGAAAAAGA ACTGATTTCT CAGTCCTTCA TTAATCTTAT TCCACACTAA	60
ATAGGTATGG GTAAACAGGT TGTTGACCTT GGTGAATCTC GACTTCAACG TCTTCGAATT	120
CTTCTACGAT TTCTTGAGCG ATTTCAATTG CAAGTTCTTC GCTTCCGTCT TCACCTACAT	180
AGAAGGTAC GATTTCACTG TCTTCATCCA ACATATGTTT CAAGGTTTCA GTCAATGTTT	240
GGTGCATATC AGGGTTTGAC ACAAGAATTT TTCCATCCAC CATACCTAAA TTATCGTTTT	300
CATGGATTTC TAAGCCATCG ATCGTTGTAT CACGCACGGC TGTGTGACG CTTCCGCTAA	360
CGACATCGCT AAGAGCAGCT GTCATACGCT CTTGGTTTTC TTCAATGGAC TTGCTTGGAT	420
CAAAGGCAAG AAGACTTGTC ATACCTTGAG GAAGAGTGCG AGCCTCTACC ACTACCGCTG	480
GTTGCTCCAA AACTTCTGCC GCAGATTGAG CTGCCATGAA GATGTTCTTG TTGTTGGCA	540
AGAAGATGAT GTTACGGGCA TTAACCTGTT CAACAGCCTT GATAAAGTCT TCTGTTGAAG	600
GGTTCATGGT TTGACCGCCT TCGATAACAT AATCCACGCC TTGAGAACAG AAGATATCTG	660

733

CTAGACCTTT ACCAGCCACC ACAGCAATCA AAGCATACTC TTTTCTTCA GCCGACTTGA	720
TAACCTTGAGT AGCTTCTTTC TCAACCTGTG CTTCTGTGTG GTTACGCATA TTGTCAACTT	780
TTACCTTGAC CAAGCTACCA TATTTGAGAC CTTCTTGCAT AACAAATCCT GGATCTTCTG	840
TATGAACATG GACTTTGACA ATTTTCATCAT CGTTAACAAC AAGGAGAGAA TCTCCAAGCT	900
CATCCAAGTA GTTACGGAAT TCATCGTAGT CAAAATCTTT AGCATAGGTT GGACCTTGCT	960
TAAGAGCTAC CATGATTTC A GTACAGTAAC CAAACGTGAT GTCCTCAGTC GCTACGTGAC	1020
CAGCTACAGA CTTATGATGC TCTACATTGA TCATCTCACT CATGTTGGCA GGAGTCGCTA	1080
CAAAGTCCTC AGATGCAATA TATTCGCCAG TAAGGGCTGA AAGGAAACCT TCGTAGATGA	1140
AGACCAATCC TTGACCACCT GAGTCCACAA CGCCAACTTC TTTCAATACT GGAAGCATGT	1200
CTGGTGTTTT AGCTAGAGCT GTTTTAGCAC CTTCCAAGGC TGGCGCATG ACTTCAACAG	1260
CGTCATCTGT TTGCTCAGCT TTTTCTTAG CACCGATAGC AGCTCCACGA GAAACTGTTA	1320
AAATCGTTCC TTCAACAGGT TTCATCACTG CCTTATAGGC AACTTCCACA CCTGATTGGA	1380
AGGCCAGAGC CAAGTCTTGA CCTGTAACT CGTCTTTATC CTTGATAGCT TGGGAAAATC	1440
CACGGAAAAG CTGAGACGTA ATCACTCCTG AGTTCCCACG CGCACCCATC AAAAGCCCTT	1500
TGGCAAGAAT GCTCGCTACT TCTCCAAGT TAGAAGCTGG CTTGTCTGCA ACTTCTTTAG	1560
CACCATTTTC AATGGTCATT CCCATATTTG TCCCAGTATC TCCATCTGGA ACTGGAAGA	1620
CGTTTAATGA ATTGACATAT TCAGCTTGCT TATTCAAGCG AGTTGATGCA GCCTGCACCA	1680
TTTCTTGAAA TAAGCTAGTA GTAATTTTG ACACGGTTAT TCTCCTACAA CTTTGATATT	1740
TTGAATGTAG ACATTTACAG TCTGAGCAGT AATCCAAGC TGGTTTTCCA AGCTAAAGGC	1800
AACACGCTCT TGAATGTTTT TTGACACTTC ACTAATCTTT GTTCCGTAGC TTAACACGGT	1860
ATATACATCA ACTGCAATAC TGCCATCTTC GGCTGCCTTT ACGACGACAC CTTTAGAATA	1920
ATTTTCCTTA CCTAGCAGGG CTGGAAATT ATCTTTGAGG GCATTTTAC TAGCCATACC	1980
GACCACACCA GAAATCTCAG TTGC	2004

## (2) INFORMATION FOR SEQ ID NO: 96:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11915 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

734

CCGGGTTGGG CTGTTGCCCC ATTAAAGCGG CACCACAGCT GGTTTCAGAA CGTCGTGAGA	60
CAGTTCGGTC CCTATCCGTC GCGGGCGTAG GAAATTTGAG AGGATCTGCT CCTAGTACGA	120
GAGGACCAGA GTGGACTTAC CGCTGGTGTA CCAGTTGTCT TGCCAAAGGC ATCGCTGGGT	180
AGCTATGTAG GGAAGGGATA AACGCTGAAA GCATCTAAGT GTGAAACCCA CCTCAAGATG	240
AGATTTCCCA TGATTATATA TCAGTAAGAG CCCTGAGAGA TGATCAGGTA GATAGGTTAG	300
AAGTGAAGT GTGGCGACAC ATGTAGCGGA CTAATACTAA TAGCTCGAGG ACTTATCCAA	360
AGTAACTGAG AATATGAAAG CGAACGGTTT TCTTAAATTG AATAGATATT CAATTTTGAG	420
TAGGTATTAC TCAGAGTTAA GTGACGATAG CCTAGGAGAT ACACCTGTAC CCATGCCGAA	480
CACAGAAGTT AAGCCCTAGA ACGCCGGAAG TAGTTGGGGG TTGCCCCCTG TGAGATAGGG	540
AAGTCGCTTA GCTCTAGGGA GTTTAGCTCA GCTGGGAGAG CATCTGCCTT ACAAGCAGAG	600
GGTCAGCGGT TCGATCCCGT TAACTCCCAT TTTAGCGGGT GTAGTTTACT GGTAAACTA	660
CAGCCTTCCA AGCTGTTGTC GCGAGTTCGA TTCTCGTCAC CCGCTTTGAA CTTTGTCTT	720
TGTACCAAGT TTTTGACTTG GGCGCGTAGC TCAGGTGGTT AGAGCGCACG CCTGATAAGC	780
GTGAGGTGCG TGGTTCGAGT CCACTCGTGC CCATAGTGT TAGTCCATTA CTAGGGGATT	840
GGAATATTAT CTGTTCACTA AGAGGACACG GGCTTGTTC CGTATAAACT ATTTTGAGG	900
ATTACCCAAG TCCGGCTGAA GGGAACGGTC TTGAAAACCG TCAGGCGTGT AAAAGCGTGC	960
GTGGGTTCGA ATCCACATC CTCCTTTTAT ATTAACGCGG GATGGAGCAG CTCGGTAGCT	1020
CGTCGGGCTC ATAACCCGAA GGTCTGAGGT TCAAATCCTG CTCCCGCAAT AAGGCTCGGT	1080
AGTTCAGTTG GTAGAGCAAT GGATTGAAGC TCCATGTGTC GGCGGTTCTGA TTCCGTCTCG	1140
CGCCATTTAT ATATTTTGA AGGGTAGCGA AGAGGCTAAA CGCGGCGGAC TGTAAATCCG	1200
CTCCTTCGGG TTCGGGGGTT CGAATCCCTC CCCTTCCATT TTACGGGCAT AGTTTAAAGG	1260
TAGAACTAAG GTCTCCAAAA CCTTCAGTGT GGGTTCAATT CCTACTGCCC GTGTAAATAG	1320
AATTATGGCG GGTGTGGTGA AGTGGTTAAC ACACCAGATT GTGGCTCTGG CATGCGTGGG	1380
TTCGATCCCC ATCACTCGCC TATTTTATAT TGGGGTATAG CCAAGCGGTA AGGCAAGGGA	1440
CTTTGACTCC CTCATGCGTT GGTTCGAATC CAGCTACCCC AGTTACTATT TGCCGGCGTG	1500
GCGGAATTGG CAGACGCGCT GGACTCAAAA TCCAGTGTCC GCAAGGACGT GCCGGTTCGA	1560
CCCCGGCCGC CCGTATAGTA TAGTGTTAGG AACGTTGTTA TTCTTCGTTT CTTTTTTATA	1620
TTATTTTGG TATAATTATA GTTATTCAAA TTTTATTAG ATTAAGAAAG TGTAGGGGAG	1680
TATGTCTTGT TCTATCGATT TATTAAAACA TCGGTATTTG AAAAATATTA AAGAAAATCC	1740
TGAATTGTTT GTCGGAATTG AGTTGGAGTA TCCTGTTGCA AGTTTAGAAG GGGATGCTAC	1800

735

AGATGTTGAA GTTATGAAGG ATCTATTTCA TTATTTAGTT TCTACTTTGG ATCTCACCGT	1860
AGCAAAGGTA GATGATTTTG GCAATCTGAT CCAGTTAGTA GATCCGATAA GTCAGGATGC	1920
TATTTTATTT GAAGTTTCCT ATACAACGAT TGAGTTTGCA TTTGGTAAGG CTGAAACGAT	1980
TCAAGAGGTC GAAAATCGTT TCAATAATTA TATGAATGTA ATTCAGAGAA AGTTAGCTGA	2040
ATCAAATCAT GCTATTGTTG GCTGTGGTAT CCATCCCAAC TGGGATAAAA ATGAGAATTG	2100
TCCAGTGGCT TATCCACGCT ATCAGATGTT GATGGATTAT TTGAATTTGA GTAGAAATAT	2160
TATTAATCA GATTTACATC ATTTCCCTGA ATATGGTACT TTTATCTGTG GGAGCCAGGT	2220
TCAGCTGGAT ATTTCAAAAA CCAACTACTT ACGGGTGATT AATGCTTTTA CTCAAATTGA	2280
AGCGGCTAAG GCTTATTTAT TTGCAAACTC TGAATTTTCG GGTGCGGATT GGGATACGAA	2340
AATTTCAAGG GATATTTTCT GGAAGAATC TATGCATGGT ATCTATCCAG AGAATGTTGG	2400
GGTCAATGCT AGACTCCTTA ATGATGAAAC TGATTTTTTT GACTATCTAA ATCATTCTGC	2460
GATTTTTACT GCGGAACGTG ATGGGCAGAC CTATTATTTT TATCCTATTC AGGCTGGGGA	2520
CTATTGGCT ACGTCCGAAA TCCAAGCATT TGCTCTGAAT GGGGATGAGG TTATTATTTA	2580
CCCCAAGAG AAGGATTTTG AAATCATCG TAGTTACCAG TACCAAGATT TAACGACTCG	2640
AGGAACAGTT GAGTTTCGTA GTGTGTGTAC ACAGCCACTT GATAGGACTT TTGCTTCTGC	2700
AGCTTTTCAC TTGGGATTAT TGGTTAATTT AGACAAGTTA GAAGCTTACT TAGAAACAGC	2760
ACCTTTCTTT AAAGTATTTG GTTATGATTA CAAGTCTTTA AGGAGACAAT TTTCTAAGAA	2820
AAATCTTACA GATGAGGAAG AAATACGAT TATTGAATTT TCCAAAGACT TACTCCTACT	2880
AGCTGAGGAG GGAAGTAGTG TGAGAAATAA GGAAGAAATG ACCTATTTAC AGCCTTTGAG	2940
AGAAGAATTG AGCCTATAAT TTCTCTTATA AAGGGAGAAT TTTCTGAAAA ATCATGATAT	3000
AATGGACGAG ACTATAGATA AAGGATAGAG AGTAATGACA TTAGTTTATC AATCAACGGC	3060
TGATGCCAAC AATACAGTAA CTGCCAGCCA AGCAATTTTG CAAGGTTTGG CGACGGACGG	3120
CGGTTTGTTC ACACCGGATA CTTATCCAAA GGTAGATTTG AACTTTGACA AATTGAAAGA	3180
TGCTTCTTAC CAGGAAGTTG CTAAGCTAGT TTTGTCAGCA TTTTATAGAT ACTTTACAGT	3240
TGAGGAGTTG GACTACTGTA TCAACAATGC CTACGATAGC AAATTTGATA CTCCAGCTAT	3300
TGCACCATTA GTGAAATTAG ATGGGCAATA CAATTTGGAA CTTTTCATG GTTCAACGAT	3360
TGCCTTTAAG GATATGGCCT TGTCTATTTT GCCATACTTT ATGACGACTG CTGCTAAGAA	3420
ACATGGTTTG GAGAACAAGA TTGTTATCTT GACAGCGACA TCTGGTGACA CGGGGAAAGC	3480
TGCTATGGCG GGGTTTGGCA ATGTGCCTGG TACTGAGATT ATCGTCTTTT ATCCAAAGGA	3540

736

TGGTGTGAGC	AAGATTCAAG	AGTTACAAAT	GACCACTCAG	ACTGGCGACA	ATACTCATGT	3600
TATTGCTATT	GATGGTAACT	TTGACGATGC	GCAAACAAAT	GTGAAGCACA	TGTTTAACGA	3660
CGTGGCTCTT	CGTGAAAAAT	TGACTACCAA	CAAGTTGCAA	TTTTCATCAG	CTAACTCTAT	3720
GAACATTGGT	CGTCTGGTGC	CACAAATTGT	TTATTATGTT	TATGCTTACG	CTCAATTGGT	3780
TAAGACTGGT	GAAATTGTAG	CTGGTGAAAA	GGTTAACTTC	ACAGTACCAA	CAGGAACTT	3840
TGGAAATATC	TTGGCTGCCT	TTTATGCCAA	ACAAATTGGT	TTGCCAGTTG	GTAAATTAAAT	3900
CTGTGCTTCA	AATGACAACA	ATGTTTGTAC	AGACTTCTTT	AAAACACGTG	TCTATGACAA	3960
AAAACGTGAG	TTTAAGGTAA	CAACCAGCCC	ATCTATGGAT	ATCTTGGTAT	CTTCAAACCTT	4020
GGAGCGCTTG	ATTTTCCATC	TTTTGGGAAA	TAATGCTGAA	AAGACAACTG	AACTTATGAA	4080
TGCCTTGAAC	ACGCAAGGAC	AATATAAGTT	GACAGACTTT	GATGCAGAGA	TTTTGGACCT	4140
CTTTGCAGCT	GAATATGCGA	CTGAGGAAGA	AACGGCAGCA	GAGATCAAGC	GTGTTTGTGA	4200
GTTAGATTCT	TATATCGAGG	ACCCTCATAC	AGCTGTTGCT	TCAGCAGTTT	ATAAAAAATA	4260
CCAATCGGCC	ACTGGAGATG	TAACTAAGAC	AGTGATTGCT	TCAACAGCTA	GTCCATACAA	4320
GTTCCAGTA	GTTGCAGTAG	AAGCTGTAAC	TGGAAAAGCA	GGTTTAACAG	ACTTTGAAGC	4380
CTTGGCTCAA	TTACATGAAA	TCTCAGGCGT	TGCAGTGCCA	CCAGCAGTTG	ATGGGCTTGA	4440
AATAGCTCCA	ATTCTGCACA	AGACAACAGT	GGCAGCTGCT	GACATGCAAG	CAGCGGTGTA	4500
GGCTTATTTA	GGACTTTAAG	ACAGAGGGAG	CAAACTCGCT	TGGGAAACCA	ACTGAGTTTC	4560
TTTTCATCAG	GAGGAGAGAT	TGTTTAAGAA	AAATAAAGAC	ATTCTTAATA	TTGCATTGCC	4620
AGCTATGGGT	GAAAACTTT	TGCAGATGCT	AATGGGAATG	GTGGACAGTT	ATTTGGTTGC	4680
TCATTTAGGA	TTGATAGCTA	TTTCAGGGGT	TTCAGTAGCT	GGTAATATTA	TCACCATTTA	4740
TCAGGCGATT	TTCATCGCTC	TGGGAGCTGC	TATTTCCAGT	GTTATTTCAA	AAAGCATAGG	4800
GCAGAAAGAC	CAGTCGAAGT	TGGCCTATCA	TGTGACTGAG	GCGTTGAAGA	TTACCTTACT	4860
ATTAAGTTTC	CTTTTAGGAT	TTTTGTCCAT	CTTCGCTGGG	AAAGAGATGA	TAGGACTTTT	4920
GGGGACGGAG	AGGGATGTAG	CTGAGAGTGG	TGGACTGTAT	CTATCTTTGG	TAGGCGGATC	4980
GATTGTTCTC	TTAGGTTTAA	TGACTAGTCT	AGGAGCCTTG	ATTCTGTCAA	CGCATAATCC	5040
ACGCTGCCT	CTCTATGTTA	GTTTTTTATC	CAATGCCTTG	AATATTCTTT	TTTCAAGTCT	5100
AGCTATTTTT	GTTCTGGATA	TGGGGATAGC	TGGTGTGCT	TGGGGGACAA	TTGTGTCTCG	5160
TTTGGTTGGT	CTTGTGATTT	TGTGGTCACA	ATTAAACTG	CCTTATGGGA	AGCCAACTTT	5220
TGGTTTAGAT	AAGGAACTGT	TGACCTTGGC	TTTACCAGCA	GCTGGAGAGC	GACTTATGAT	5280
GAGGGCTGGA	GATGTAGTGA	TCATTGCCTT	GGTCGTTTCT	TTTGGGACGG	AGGCAGTTGC	5340



737

TGGAATGCA ATCGGAGAAG TCTTGACCCA GTTAACTAT ATGCCTGCCT TTGGCGTCGC	5400
TACGGCAACG GTCATGCTGT TGGCCCGAGC AGTTGGAGAG GATGATTGGA AAAGAGTTGC	5460
TAGTTTGAGT AAACAAACCT TTTGGCTTTC TCTGTTCCCTC ATGTTGCCCC TGTCCTTTAG	5520
TATATATGTC TTGGGTGTAC CATTAECTCA TCTCTATACG ACTGATTCTC TAGCGGTGGA	5580
GGCTAGTGTT CTAGTGACAC TGTTTTCACT ACTTGGGACC CCTATGACGA CAGGAACAGT	5640
CATCTATACG GCAGTCTGGC AGGGATTAGG AAATGCACGC CTCCTTTTT ATGCGACAAG	5700
TATAGGAATG TGGTGTATCC GCATTGGGAC AGGATATCTG ATGGGGATTG TGCTTGGTTG	5760
GGCCTTGCCCT GGTATTTGGG CAGGGTCTCT CTTGGATAAT GGTTTTCGCT GGTATTTCT	5820
ACGCTATCGT TACCAGCGCT ATATGAGCTT GAAAGGATAG GAAATGCAAA AAACAGCTTT	5880
TATTTGGGAT TTAGACGGGA CTTTATTGGA CTCTTACGAA GCGATTTTAT CAGGGATTGA	5940
GGAGACTTTT GCTCAGTTT CTATTCCTTA TGATAAGGAG AAGGTGAGAG AGTTTATCTT	6000
CAAGTATTCG GTGCAAGATT TGCTTGTGCG GGTGGCAGAA GATAGAAATC TGGATGTTGA	6060
GGTGCTAAAT CAGGTGCGTG CCCAGAGTCT GGCTGAGAAG AATGCTCAGG TAGTTTGTAT	6120
GCCAGGTGCG CGTGAGGTGC TAGCTTGGGC AGACGAATCA GGAATTCAGC AGTTTATATA	6180
TACTCATAAG GGAACAACG CTTTTACCAT TCTCAAGGAC TTGGGGGTGG AATCCTATTT	6240
TACAGAGATT TTAACCAGTC AGAGTGGCTT TGTGCGGAAG CCAAGTCCAG AAGCGGCTAC	6300
CTATCTGCTA GATAAGTATC AGTTGAATTC TGATAATACT TATTATATAG GGGATCGGAC	6360
TCTGGATGTG GAATTTGCCC AGAATAGTGG GATTCAAAGC ATCAACTTTT TAGAGTCTAC	6420
TTATGAAGGG AATCACAGGA TTCAAGCGTT AGCAGATATT TCCCGTATTT TTGAGACTAA	6480
GTGATAAAAA GATTGTGTCA GTTTTGTGAC AGAGACCTAA CAACTATTT CAAGTAACCT	6540
AGTTTGTTAC AAGGAATAGA CAGTTCTGTT AAATAGGCCC GAGAGGGCTT TTTTCTACA	6600
TTTTTTGTGT TATGATAGAC AGGTACTCAT TTGAAAGGAA TTTGAAAGAA TGAAGAAAAG	6660
AATGTTATTA GCGTCAACAG TAGCCTTGTC ATTTGCCCA GTATTGGCAA CTCAGCAGA	6720
AGAAGTTCTT TGGACTGCAC GTAGTGTGTA GCAAATCCAA AACGATTTGA CTAACCGGA	6780
CAACAAAACA AGTTATACCG TACAGTATGG TGATACTTTG AGCACCATTG CAGAAGCCTT	6840
GGGTGTAGAT GTCACAGTGC TTGCGAATCT GAACAAAATC ACTAATATGG ACTTGATTTT	6900
CCCAGAAACT GTTTTGACAA CGACTGTCAA TGAAGCAGAA GAAGTAACAG AAGTTGAAAT	6960
CCAAACACCT CAAGCAGACT CTAGTGAAGA AGTGACAACT GCGACAGCAG ATTTGACCAC	7020
TAATCAAGTG ACCGTTGATG ATCAAACTGT TCAGGTTGCA GACCTTTCTC AACCAATTGC	7080

738

AGAAGTTACA AAGACAGTGA TTGCTTCTGA AGAAGTGGCA CCATCTACGG GCACTTCTGT	7140
CCCAGAGGAG CAAACGACCG AAACAACCTCG CCCAGTTGAA GAAGCAACTC CTCAGGAAAC	7200
GACTCCAGCT GAGAAGCAGG AAACACAAGC AAGCCCTCAA GCTGCATCAG CAGTGGAAGT	7260
AACTACAACA AGTTCAGAAG CAAAAGAAGT AGCATCATCA AATGGAGCTA CAGCAGCAGT	7320
TTCTACTTAT CAACCAGAAG AGACGAAAAT AATTTCACA ACTTACGAGG CTCCAGCTGC	7380
GCCCGATTAT GCTGGACTTG CAGTAGCAAA ATCTGAAAAT GCAGGTCTTC AACCACAAAC	7440
AGCTGCCTTT AAAGAAGAAA TTGCTAACTT GTTTGGCATT ACATCCTTTA GTGGTTATCG	7500
TCCAGGAGAC AGTGGAGATC ACGGAAAAGG TTTGGCTATC GACTTTATGG TACCAGAACG	7560
TTCAGAATTA GGGGATAAGA TTGCGGAATA TGCTATTCAA AATATGGCCA GCCGTGGCAT	7620
TAGTTACATC ATCTGGAAAC AACGTTTCTA TGCTCCATTC GATAGCAAAT ATGGGCCAGC	7680
TAACACTTGG AACCCAATGC CAGACCGTGG TAGTGTGACA GAAAATCACT ATGATCACGT	7740
TCACGTTTCA ATGAATGGAT AAACCCGACT TGATAACATC ATTTTGACGA ATGAGATCTA	7800
GCTTTCGTGA TGGAAAGCGA TTCTCGTTCG TTTTCTCTT GTCATACTCT TCGAAAATCT	7860
CTTCAAACCA CGTCAGTTTT ATCTGAAACT TCAAAGCTGT GCTTTGAGCA ACCTGCGACT	7920
AGCTTCCTAG TTTGCTTTTT GATTTTCATT GAGTATCAAT TTGAATGGAA AATGGAAAGT	7980
TATCATCTTG TAATGAGTTA AGCAACATTC TTGCAATCTA TTTTACTTTA TATCACAATT	8040
AATTAGTCAA ATATTGATAA ATCAATAAAA AGAGAGGGGA AGAAATGCTA GAGATTCAAG	8100
ATTTACTGTA TCAACTCCCG TTGTCTGAGC AAGCGAGTAC GCAATTGTTT GAAAAAAGGC	8160
TTGGGATTAG TTTGACACGG TATCAGATTT TACTGTTTTT GCTGGAGCAT TCTCCTTGTA	8220
ACCAAATGGC GGTTCAGGAG CGTTTGAAAA TTGATCAGGC TGCTTTGACA CGGCATTTCA	8280
AAATTTTGGA AACGGAAGGT TTGGTGGAGC GTCATCGTAA TCCTGAAAAT CAGCGGGAAG	8340
TGTTGGTAGA GGCTGCGAAG TATGCCAAGG AGCAGTTAGT GGTGAATCCC CCTCTGCAAC	8400
ATATCAGGGT TAAGGAAGAG ATAGAAAGTA TCTTAACAGA GTTTGAGAGA ACAGAACTCA	8460
GCCGTTTATT AAATAAATTG GTTTTGGGTA TTGAAAATAT AGAAATTTAA GGAGAAATAG	8520
ATGTCAATTA TTTAACAAC GATCGTTGCT TTGGAGCATT TTTACATTTT TTATTTGGAA	8580
AGTATTGCCA CGCAATCAGA TGCGACTAGT CGTGTATTTA ATATGGAAAA GGAAGAATTG	8640
GCTCATCCGT CAGTAAGTTC ATTGTTCAAA AATCAAGGAA TTTATAAGGC TCTGCTAGGA	8700
GTCTTTCTCT TGTATGTCAT TTATTTCTCA CAGAAATTTAG AAATGTGAC TATTTTGTG	8760
TTATTTGTGA TTGGTGCTGC GACTTACGGC TCTTTAACAG CGGATAAAAA AATTATTTTG	8820
AAACAAGGTG GATCAGCTAT TTTGGCCTTG ATTAGTATTT TACTCTTTAA ATACACTTGA	8880

739

AGGTCGATTC TAATCTCGCT AATCCTTTTT AATCCAGAAT AAGGGAAATA TGTTATACTT	8940
GTTTTTAAGA AAAAAGTCTC ATTGAATTGG TTTTGAGGAG TTAGAAATGA AAGTATTAGT	9000
GACAGGTTTT GAGCCCTTTG GAGGGGAAAA GGGCAATCCA GCTTTGGAGG CCATTAAAGG	9060
TTTACCAGCT GAAATCCATG GTGCTGAGGT CCGTTGGCTA GAGGTGCCGA CAGTTTTTCA	9120
CAATCTGCT CAAGTATTGG AAGAAGAGAT GAATCGTTAT CAACCTGACT TTGTCCTTTG	9180
TATTGGGCAA GCTGGTGGA GAACTAGTTT GACACCTGAA CGAGTGACCA TTAATCAAGA	9240
CGATGCATGC ATTTCTGATA ACGAAGATAA TCAACCGATT GACCGTCCCA TTCGCCCAGA	9300
TGGTGCTTCG GCCTACTTTA GTAGTTTGCC GATTAAAGCG ATGGTTCAAG CTATAAAAAA	9360
AGAGGGCTTA CCGGCCTCTG TTTCCAATAC GGCAGGGACT TTTGTCTGCA GCCATTTGAT	9420
GTATCAGGCT CTCTATTTGG TAGAAAAGAA ATCTCCATAT GTTAAGGCAG GTTTTATGCA	9480
TATTCCTTAT ATGATGGAAC AGGTGGTGAA CAGACCGACT ACTCCAGCTA TGAGTTTAGT	9540
GGATATTCGG CGAGGGATAG AAGCAGCAAT CGGCGCTATA ATAGAACATG GAGATCAGGA	9600
ACTCAAGTTG GTAGGCGGAG AAACCTATTG ATAGAAAAA GCTTGAGGGG AAAACCTTC	9660
AAGCTTTTGG ACGTTTTCGG GCCAATACTG CTCGGTAAAA CATAATTTTA GTGCATTGGA	9720
TATAAGGTAG GAGTGAAAAA CTAGCAATGC CAAAGGTAAT CCAATTGAGG AAGTACCAAG	9780
GAAGAAGCTG TAAATCTAGG ACAAAGTGCT GGAAGTTGTA GCCCTTCATA AAGGAACGGC	9840
TAGTTTTTAG GATTCGTCTT GGTGGGACCT GTCCTAGGTC TAGACTATAA CAGAGAAGAA	9900
ATTCCACCTG TGAATAGGCA TAATACTGTG GAATATAGAG GATATTTCCT ACAATGATCA	9960
AGATGAGACT TGCAAGAAAG TAGAGTCCAA AGACCATGAG GAAACGCTCG GTTCAACTC	10020
ATGAGAGATC TAGATTTGGA AACTCAGGAT GTAGGGTGAC GAATTTTTTG GCTAAAAAGC	10080
TACTATAAAA GAGGAGGTAA ATCCCAAGTA AATTAGGCAT ACTCCATAAA AAGAGATAGA	10140
AACGTTTGAG AAGTAGGGTC AAAAAGGTTT GAGAAAAGCG CTCCTCATCA AAGAGAGCTA	10200
GGCTGTTTTT TACAGATGGC TCCGTTTTAG AATCTTTCAT GAGTGTCACT GTTGCATAGA	10260
CGGAACTGGT CAAAAGAATA GTCCCGATAA AGGAGACTAG TAGAGGAAAG AGGTAGGTTT	10320
GAAGTATTTG GCCAAGTATG CTGAAAAATG GCTGTTCTAA AACAGTCCCG TGGATCCGAG	10380
ATAAGGGATT AAGAAAACCA GATAAGATGA CCAGCATACT GGAAGGATA TAGAGGAGAA	10440
AGAGACGGGG GGTGTCAGCC TGAAAATGTT TTGACTCCTG ACGAATTGTT TTAAATCAA	10500
TTTTTGATA GTTCATTCTC TTATTATACC ATAGTTCTTA TACATAGTTC GTGACAGTTC	10560
CTACTTTTTT TGATAAAATC ATACAGTGTG TCCTTGGGCA CACTGTATGA ACTGGGACTG	10620

740

TCTTTCCAG CTTCGGAGGT AAAAAATGTC AGATTACCA ATCAAATATC GTTTGATTAA 10680  
 GAAAGAAAA CACACAGGAG CTCGTCTGGG AGAAATCATC ACTCCCCACG GTACCTTTCC 10740  
 GACACCTATG TTTATGCCAG TTGGGACACA AGCCACTGTC AAAACTCAGT CACCTGAAGA 10800  
 ATTGAAGGAG ATGGGTTCGG GAATTATCCT ATCAAACACC TATCATCTCT GGCTTCGCCC 10860  
 TGGAGATGAA CTCATTGCAC GCGCTGGTGG TCTCCACAAG TTCATGAATT GGGACCAGCC 10920  
 TATCTTGACA GATAGTGGTG GTTTTCAGGT TTATTCTTTA GCAGATAGCC GTAATATCAC 10980  
 AGAAGAAGGA GTAACCTTTA AAAATCATCT AAATGGTTCT AAGATGTTCC TATCCCCAGA 11040  
 AAAAGCCATC TCTATTGAGA ATAATCTGGG TTCAGACATC ATGATGTCTT TTGATGAATG 11100  
 TCCTCAGTTT TATCAACCTT ATGACTACGT TAAGAAATCG ATCGAGCGTA CCAGCCGTTG 11160  
 GGCTGAGCGT GGTTTGAAGG CTCACCGTCG TCCACATGAC CAAGGTTTGT TTGGAATTGT 11220  
 GCAAGGTGCA GGATTTGAAG ACCTTCGCCC CCAATCAGCT CATGATCTTG TCAGCATGGA 11280  
 TTTCTCAGGC TACTCTATCG GTGGTTTGGC AGTGGGAGAA ACCCATGAAG AGATGAATGC 11340  
 GGTCTTGGAC TTTACAACCT AACTGCTGCC TGAAATATA CCTCGTTATC TGATGGGTGT 11400  
 GGGAGCGCCA GATAGCTTGA TCGATGGGGT CATTCTGGG GTGGATATGT TTGACTGTGT 11460  
 CTTACCGACT CGAATTGCTC GTAACGGGAC TTGTATGACC AGTCAAGGAC GTTTGGTTGT 11520  
 GAAAAATGCC CAGTTTGCTG AGGACTTTAC GCCACTGGAT CCTGAGTGTG ATTGCTACAC 11580  
 ATGTAATAAC TATACACGGC CTTACCTTCG TCACCTGCTC AAGGCTGATG AAACCTTTGG 11640  
 TATCCGCTTG ACTAGCTACC ACAATCTTTA CTTCTTGCTT AACCTGATGA AGCAACTGCG 11700  
 ACAAGCCATC ATGGATGACA ATCTCTTGA ATTCCGTGAG TATTTTGTGG AAAAAATG 11760  
 CTATAATAAG TCAGGACGTA ATTTCTAAAA TGGAAATTGAT ATAAAAAAT CCTAAGTTTT 11820  
 CTCTTAGGAT TTTTCTTCTT TTTTGTAGT AATAAAGTGT ACAATGAAAG GAAGAATAAA 11880  
 CTCGTATGCG CATTAAATGG TTTTCCTCGA TTAGG 11915

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9069 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

GAGAGGGCAA CAGTTCTATC GCTTCAAATT TTTTCTTGGT TTGCAGATAT TCAAGAATCG 60  
 GGAGTTTTTC TATAGTATTC GGCAGATTTA TTACAGCCAA GCATCTCAAA AATACGGACA 120

741

GCATCCTCCA	TCTTTTCTG	GCCTTCCTTG	ACTCTACCTT	GCTTGCTATC	AAGGAGACCT	180
TCTGCCCACA	GATAACAAT	TCGGAAATAG	GTCTCATTTT	CCTTGTAGAA	ATGCTCTTCG	240
ATAACACGTT	TAAAATAATA	GGCATTGGTA	AATTCTTCAC	ACTCAATACT	AGCTAAAAAG	300
CCATTCAATA	GTATAGTATG	AAAAAGGTTT	CGATTGCCAG	ACATTTCCAT	TAGAAAATCA	360
GATTTACGTA	CCATTTCTCG	TACATATCTA	GTAAAAAGAG	AAACAGATAA	AAATGGAGAA	420
CTGACTGAAA	ATAAATGAG	TTCATAGATT	CCCCAGATCT	CGGTAGAAAA	CAAATAATCA	480
TGAAGGACTT	TCCTTCCTC	TGCTGTTAAG	TCTACCTTTT	CATCTATGCT	CTTCATATAA	540
GACTTGATAA	TAATGGCATT	TAGAATATGT	TTCTGTTTGT	TGTGAGAAATG	GGCATGCTTT	600
TATACTCCCT	GCGATATAAG	TCCTCAAGAG	GTGCTATATT	CTTTGGTTCC	AAGACATCTG	660
TAATTTCTTT	TCTCAACTCA	GAATCTGTAT	CATACTGGAA	ACCTCTTGCC	AGAAAGAGGA	720
TCTCCTCCAC	ACTGGCAGAT	ATATTTTCCA	GAGCAAATAG	AAACTTTTCC	ACCGAAAGCT	780
CACCTCTGACC	TGTTTCAAAA	CGGGACAACA	TAGACGGCGA	AAATTGTCCT	CCGGTTGCTT	840
GTCTCAGTGA	GATATTCTT	GACTCTCGTA	ATTGTCTAAA	GACTTTTCCA	ATCTGCTCCA	900
TAGACTTCCC	CTTGATTCCG	TATTTTCTTC	ATTTTATCAT	ATTTTTCAGA	AAATTCATCA	960
AAAACTTGCC	AAATTGTGAG	AATTATGAGA	AAATAGAGGA	TATTTATCAC	GTGGAGGGAC	1020
TGCTATGAGA	GACGATATCA	AAATCAATGA	CCGTGCTTTG	GCCTTGCAAG	ACCAAATTAT	1080
CGAAAACTA	GAGAAAGTTT	TTGATACAGA	TGTGGAATTG	GATGTTTACA	ATCTAGGTCT	1140
GATTTATGAA	ATCAATCTGG	ATGAAACGGG	GCTCTGCAAG	ATTGTCATGA	CCTTCACCGA	1200
TACTGCCTGT	GATTGCGCCG	AAAGCCTGCC	TATTGAAATC	GTGGCAGGTC	TGAAACAAAT	1260
CGAGGGTATC	AAAGATATCA	AGGTTGAAGT	TACCTGGTCG	CCTGCTTGGA	AAATCACACG	1320
AATCAGTCGC	TATGGCCGTA	TTGCCCTTGG	ACTACCACCT	CGTTAAGCAG	ACCAATCACT	1380
TTTAAAGATG	AAAATCAAAG	GGCAAAC TAG	AAACTAGCC	GCAGGTTGCT	CAAAACACTG	1440
TTTTGAAGTT	ATGGATAGAA	CTGACGAAGT	CAGCTCAAAA	CACTGTTTTG	AGGTTGTGGA	1500
TAGAACTGAC	GAAGTCAGCT	CAAAACACTG	TTTTGAGGTT	GTGGATAGAA	CTGACGAAGT	1560
CAGCCCCAAA	CACTGTTTTG	AGGTTGTGGA	TAGAACTGAC	GAAGTCAGTA	ACCATACCTA	1620
CGGCAAGGCG	ACGTTGACGT	GATTTGAAGA	GATTTTCGAG	TATGAGTTTA	TTTTTTACCT	1680
GACTTGTCCA	TATCCAGAA	GTCTGTCACG	GCTCCGCGTG	AAGCAGATGA	TACGATGTGG	1740
GCATATTTAC	CGAGGACACC	ACGGCTGTAA	AGTGGTGGCA	AGGTTGTTTC	TGCCTTGCGT	1800
TTTCAAGTT	CTTCTTCGGA	TACGGCCATA	GAAATTTCTT	TGGTATCTTG	GTCAACCGTA	1860

742

ACGATATCGC CGGTACGGAG ATAGGCAATT GGTCCACCAT CCTGAGCTTC AGGAGCGATA	1920
TGTCCAACAA CCAGACCATA AGTACCACCA GAGAAACGTC CGTCCGTCAA GAGGGCCACC	1980
TTATCTCCCT GACCTTTACC AACAAATCATT GAAGAAAGTG ATAGCATCTC AGGCATACCA	2040
GGACCACCTT TAGGTCCAAC AAAACGAACA ACGACTACAT CGCCATCAAC GATTTCATCT	2100
GTCAGAACGG CCTGAATCGC ATCTTCTTCT GAGTCAAAGA CCTTAGCTGG CCCAACGTGA	2160
CGACGCACTT TAACACCTGA TACCTTGCCA ACTGCACCGT CAGGAGCAAG GTTCCCGTTC	2220
AAGATGATAA GCGGACCATC CGCACGTTTT GGATTTTCAA GTGGCATGAT AACTTTTTTG	2280
CCTGGAGTCA AGTCTGCAAA GTCAGCCAAG TTTTCAGCTA CAGTCTTACC AGTACATGTG	2340
ATGCGATCTC CGTGAAGGAA ACCATTTGCC AACAAATACT TCATAACCGC AGGGACACCA	2400
CCGACTTCGT AGAGGTCTTG GAAGACATAC TGACCAGATG GTTTCAGTC GGCCAAGTGA	2460
GGCACACGTT CTTGAATCGT ATTGAAGTCC TCAAGTGACA AGTCAACATT TCGGGCATGG	2520
GCAATGGCGA GCAAGTGAAG AGTGGCGTTT GTAGAACCAC CGAGAGCCAT CGTTACAGTG	2580
ATAGCATCTT CAAAGGCTTC ACGAGTCAAG ATATCTGATG GTTTGAGACC AAGTTCCAAC	2640
ATCTTAACAA CAGCACGTCC TGCTGCTTCG ATATCTTCTT TCTTATCAGC TGATTACAGT	2700
GGGTGAGAGG ATGACCCTGG CAAACTCATC CCTAGAACTT CGATAGCAGT TGCCATGGTA	2760
TTAGCAGTAT ACATACCACC ACAACCACCA GGGCCAGGGC AGGCATTACA TTCAAGACGT	2820
TTCACGTCCT CAGCTGTCAT GTCACCGTGG TTCCATTTTC CGATACCTTC AAAGACAGAA	2880
ACCAAGTCGA TATCTTTACC ATCAAGATTT CCCGGTGCAA TAGTTCCACC ATAGGCGAAA	2940
ATAGCTGGGA TATCCATATT AGCAATAGCA ATCATAGATC CAGGCATGTT CTTGTCACAG	3000
CCACCGATAG CGACGAAGGC ATCCACGTTG TGACCACTCA TAGCCGCCTC GATGGAGTCC	3060
GCGATGATGT CACGAGATGT TAGAGAGAAA CGCATACCAG GCGTTCCCAT AGCGATCCCG	3120
TCCGCTACGG TAATGGTTCC AAAGTGTACA GGCCAAGCGC CTGCAGATTT GACACCTTCT	3180
TTAGCCAGTT TCCCGAAATC ATGCAAGTGA ATGTTACATG GTGTATTTTC CGCCCAAGTC	3240
GAAATCACTC CCACAATCGA TGTTTCAAAG TCCTTATCTG TCATACCAGT CGCACGAAGC	3300
ATAGCACGGT TAGGTGATTT AACCATGCTG TCATAAATGC TACTGCGGTG ACGTTTATCT	3360
AATTCAGTCA TCTTATCCCT CCCATTTTCTG TTTTACTAT TATAGCACAA TTTTCGCATG	3420
AAGAACAGAA TAAAATCTT GAATTTTCTG AAAATTTCTAT ACACATGTGA AATATTTAAA	3480
ATTAAAAACA ACAAAGCGGA TTAGTGCACCT TTCTGATGAC CAGAATATGC TTTTAAATCC	3540
GCTTTCTTTA AATAACGTAC TGTAATTTTT ACAGAAATTC TTTCAAATAA GTGTATTTAA	3600
CATCTATCTT GCATTATAAA TTTCTAGAAC CTTCTCTTTT ATATTGATT CACTCAAACC	3660

743

ATACTCATT	AGAAGATAAT	CCATTTTCCC	TACTTGACCG	AATCTTTCTT	GAACACCCAT	3720
CCGATGAAT	TTTGTTATTC	CATCATCAGA	GAATAATTCA	CATAAAGCAC	TGCCAATTCC	3780
ACCTATCTGA	TTGTGGTTTT	CTACAGTAAA	TATAGTTTTT	CCACTTAACA	TTGTTTTTAT	3840
CTGTTCTGGT	ATCGGTTTGA	TTCTAAATAA	ATCTATCACA	CCTACTGAAT	AACCTAATTT	3900
AGACAGTTCA	TCTGCAACTC	GAATACTTGG	AGCAACCATT	ATGCCAGAAG	CAACGATTAC	3960
AAGATCTTCA	CCATGCCTTA	ACTCAATGTA	GCCTTTAGAA	AAATCTTCTC	CACCTTGATA	4020
CACAGGAACT	GGAGCTTTTC	TAATTGTTTCG	AATATATTTT	AGTCCTTTTA	AGTCTAATGT	4080
CTGGTTCAAT	ATTTCACGAA	ATTGGATATC	ATCAGTTGCT	TCGAAAATGA	TTGATTTAGG	4140
AATTAAACGT	AACAATCCAA	TTTCTTCAAA	TGGCATATGT	GTTCCACCAT	TCATCTCTGC	4200
CGTTACTCCT	GCATCTGATC	CAATCACAGT	GGCATCCAAT	TGTGCGTATC	CAAGAGAAAT	4260
AAATAATTGA	TCAAATACTC	TTCTGTGAAGC	AAAAGGACCA	AATGTATGAA	GATAAGGTCT	4320
AAACCCCTGA	ATAGACAAGC	CTGCTGCAAG	GGCGACCATT	TCTGCTTCCA	TAATCCCAAC	4380
ATTACATATA	CGGTCTCCAA	AGTCCTTTTC	AAGATTATTA	GTAGCCATCG	AACTTGACAA	4440
ATCGGCTTCT	AAGACTACTA	TATCAGAATC	ACTTTGATTA	GCCTCTAAAA	GGAAGTCTCT	4500
ATATACATGC	CGTAATTCTT	TGCTACTTCT	CATCATCTCG	TTTCTCTCAA	TTCTTGACTT	4560
AATCTTTCTA	CAACTGAAGT	TAACATTTGT	TTCTCCTCTA	CAGTAGGGCG	AAGATGATGA	4620
TTGGATTTCA	TTTCTTCCAG	CTCTTGAACC	CCTTGACCTT	TAATAGTATC	TAATACAATG	4680
CACCTAGGTG	ATGAATTATT	TGACTGTTTT	AATTGGACAA	TCCCTTCATA	AATTTCTCTA	4740
ATATCTGAAC	CCTTGACCCT	AATGGATTCA	AATCCAAATG	CTGAAAATTT	TTCTACGAAA	4800
TCACCTGGAT	TACAAATATC	CTTTGTAAAA	CCATCTAATT	GTTTTTTGTT	ATCATCAACA	4860
AATACAATTA	AGTTGGATAA	CTGTTGATGA	GAAGCAAAC	GTATAGCCTC	CCAACATTGT	4920
CCCTCATTTA	ACTCACCATC	TCCAACAATA	GCGTAAGTAT	AAAAGGGACT	CTTTCTTATT	4980
CTCTGACCAT	ATGCAAGTCC	AGTTGCAACA	CTAATTCCTT	GTCCTAAAGA	GCCCGTTGTC	5040
ATATCTATGC	CTGGCGTTAG	ATTTCTATCA	GGATGAGACG	GTAATTTGGT	TCCATTTGTA	5100
TTTAAAGAAT	ATAAGAATTC	TTTGTCAAAG	AAACCATTCA	AATAGAGTGT	ACTGTATAGA	5160
GCTGGTCCTC	CGTGACCTTT	TGATAATATG	AAATAATCTC	TATCTCGTGC	TGCAAATATT	5220
TCTGGAGTCA	TTGGCATTAT	TTCAACATAA	AGCACCGCTA	AAACTTCTAC	GATAGACAGA	5280
CTTCCTCCGT	AATGTCCGAA	TCCAAGATGA	TTCAATGTTT	TAAGAGTATT	TAATCGGATG	5340
TTAGTCGCAA	ATTTTCTTAA	CCCATCTTCT	CTATTTTAC	TTAAAATCAT	CCCTTATTCC	5400

744

TCCGTTGCAG ATGGCTTTTT AATAAAGGAT ACTCCAAACA TAACTGCTAG AATAAGAACA	5460
AGACCAATCA CAATGCCTGC TTGTGAGCCA AATTGATTTA ACATTCCTAA AATAATTCCT	5520
GATAGACCAA AATCTGCATC TGAGAAAGTT GATCCTTGGA AACCAAGTCC TCCCAAACT	5580
GGCATTAAAA AGACTGGAAG AAAACTGATT AAAATACCTT GTAAAAATGC TCCAATAGTG	5640
GCTCCACGAA CACCACCAGA TGCATTCCCA ATGACACCTG CAGTCGCTCC ACAGAAGAAA	5700
TGAGGCACAA CACCTGGTAA GATAACAACC GTTCTGAAG CAATCATAAT TACCATACTT	5760
ACTAAACCAC CAACAAAAC AGAGATAAAT CCAATTAGAA CTGCATTGGG TGCATAAGTA	5820
TAAACAATCG GACAATCCAA AGCAGGTTTT GAATTAGGTA CAAGACGCTC TGAATACCT	5880
TTAAAGGCTG GAACAATTC GCCCAAAATA AGGCGAACAC CTGCTAAAAT AACAAATACC	5940
CCTGCTGCAA ATTGACCTGC TAATTGTAAA GCATAAACTA GACCACTTGT ACCACTACTG	6000
ATTTCTTTTT CTATATATTC TGACCCTGCA AAGATAGCTA CAATAATGTA AATAACTGCC	6060
ATGGATAAAG TAATACTAAC AGTACTATCA CGTAAAAAG CTAACTCTT TGGAAATTTA	6120
ATGTCTCTG TTGATTTGA TTTGTCACCG ATAAGGCTAC CAGTAAAACC ACTCAACCAA	6180
TATCCCAAAG AACTGAAATG ACCTAAAGCT ACCTTGTCAT TTCCAGTTAA TTGAACCATA	6240
TATTTTGGCA CAAATGCTGG GGAAATACTC ATAATAATAC CGAGTGCTAA TCCTCCTAGT	6300
AAGATGAGAG GCAAGCTAGT AAAGCCAGCA ACTGATAAAA TGACCGCAAT CATACTGCC	6360
ATATATAGAG TGTGGTGCCC TGTAAAAAA ATATATTTAA ATCGAGTAAA ACGAGCGATT	6420
AAGATATTGA ACACCATGCC TGCAAACATA ATCATTGCAG TAGCTGAGCC ATATGTTGTT	6480
AAAGCTACAG CTACAATTGC TTCATTATTC GGCACAACGC CAGATAAATG AAAAGCATGC	6540
TCAAACATGG TACCAAATGG ATTTAAAGAA TTTTGACAA TTCCTGCACC ACCAGATACA	6600
ACTAAGAAAC CAACAAAGGT CTTAATTCCA CCTTTAATAA TATCAGGTAA TTTCTTCTTC	6660
TGAAGAACTA ATCCTAAGAT TGCAATTAAA GCTACTAAAA TAGCTGGTGT ACTAACAATA	6720
TCCAATATGA ACTTCATCAT GACGCTAGCC TCCTATATAA GTCCTTTTTC TTCACAAAT	6780
TTAGTAATTA ATTCTCGTAG TTCATCCATA TCAATAATAC TATTTAAGAT ACGAACATCT	6840
CCAAGATGAC TAGCTGAATC AGCTAGATCA CGACCAACAA TCCAAATATC AGCTGCATTT	6900
GGATCTGCTC CACCTAAATC ATAATGTTCA ACTTCTACAT CCGAAACATT CAAATCACTC	6960
AATACAGATT CAATATTCAT CTGTACCATA AAACCTGAAC CTAATCCTGA ACCACAAGCT	7020
GTACCAATTT TTAACATTAT CTAATCCTCC TGTMTAATTA TCATTTTAAT GTCATCATAG	7080
TTTTTTGATG ATATTAAAGT TTGAACATGA TTTTATCTC TTAAAAATTGT TGTAAATGT	7140
GACAAAGCCT TTAAATGACT CTCATTATCA ATGGCTGCAA TACAAATCAA CAATCTTACC	7200



745

TCTTGTTCTG GATTATCCAA TAAATAAATC GGTTCTTCCA AAAC TAACAT TGACATTCCCT	7260
ATTTCATTCA CACCTTCATC TGGCCGAGCG TGAGGAATTG CTACTCCCTT CCCTAAATTA	7320
ATAAAAGGTC CAAACTCTTC TACTTTTGA ATCATTGCCT CAGGGTAGTT CTCAGTTATC	7380
TTATCTTGAT CAAAAGCGG TTTAGCTGCT AAACGAATCG CCTCCTTCCA TCCTAATTTT	7440
TGCGAACTAA CCTGATAGGT TTCTTTGGTA ATAAGTTGTT CTAGCACTGG TACAATTTCC	7500
TTTCTATCAT TTTTTGGTA AAGATAATTC TTAAACGCCA ATCTTAATTC CAATCTTGT	7560
GTAATAATTC CATATCTTTT GACAATATTC AGGATTGTT CAATCTCAAA ATCTCCATAC	7620
TCTAAATTCG GAAATCTTT TAACACTAGT TCTACTAGTT GTATTGCTTG CTCTTCAGTC	7680
ATCATAACCG AAAC TAGATA ATTTGGCTTT TCTGTCTCCA CCTTTATGGT AGAAAAAACC	7740
ATATCATAGT CACTACTAGC TTTCACCTGT AAATCATCAA TCTTTGAGGT TCCTATAAAC	7800
TCAATTTGAG GAAATAATGC TAATAGATTC TCTTTTAACA TCAATGAAGA ACTAACACCA	7860
TTAGGACAAA TGATTGCTGC TTTATACCAT TTTGAGGCA AAGTATCTGC TTTCTTAAA	7920
TAACCTCCGA AATGGATAAC AAAATATGCT GTTTCACTAT CAGGTATGGG ATTGTCAATA	7980
CGTCCATCA AGGGCATCAA AGAATCTTGT ACTAATTCOA ATAAATCAGG ATAATGTTCT	8040
TTAATGCA ATACATATTC ATTTGAACTA GGTAGGCCGA ACTTTAATCT ATAGTAAGCC	8100
GGTATAAGGT GGCGCGAAG ATTTTCTCTC AATCCTTCCC TTTGTTTAAA ATGTAACAAA	8160
GAAATATCTT CCATTCTACT TATAATAGCC TCTGTTAATT GATTAAAGTA AACCGGAGCA	8220
ACATCTACTT CACCTTCAAA GCAACTTGAT AATAAACCGG TGATATAGCG ATAATCATCC	8280
TCAGAAAACA CCGTATCTAT AATTCCCAAA TCAACCACTG TATCCAATAA AATAGTGGTT	8340
ATATCTTGAA TAACAGGAGA TACTAATGTC TCTGAAAGAC ATACTCTTTC AACATCCCTT	8400
TGATACCTAC ACAGAATGAA TACTAAACCG AAAAGGTAAA CTTTTAATTG ATTAACAATA	8460
GGTACTAGCT GTAGCTTCTC ATAATAATCT TTAACCTACCT GATCAATCAA ATCATAAGTT	8520
AATGAATACC CCCAACTGGA TAAAACATAA TCCAAACCCC AAATCCCTAT GGAGGATTCC	8580
AGCAACTCAC TAACCATTTG AAAAGCTAAG CGGTGCTTAT TCCACTCTGA ACCGTGTAAA	8640
GTATAACCTT TTGCTCTACT GTACCCTAGC TCCAAATCAT TATCTAACAT AATCTTTCTT	8700
AATGATTGAA TATCAGATAA GGTGTATTTC TTACTTACTT TCAAAAAGTC TTGGTAATGA	8760
CTATTCGATA TAAATCTAA TCGGCAAAAA GTGTAAAGAT AGATTAAAGC TAAGCGAGTC	8820
GACTTTGGTA AAACCAATTC ATCCGACTTA ATAATATCTG TCAAAGACTG CTTCGTACGA	8880
TTTGATAAAC TATAGCGACC TTGCTTTTAA TCCAGCACTA TCCCTTTATT AGCTAGATAA	8940

746

GGCACTAAAT AATCTATTCC TTCTTTGACT TCCTTTATAG GTAAGCTCAC CTTAACAGAT	9000
AATTCATATA ACGATAGCTC ACAATGATCC ATCAAAGTCA TCAAAATAAC TAGTGCTCTA	9060
TAATCAAAC	9069

## (2) INFORMATION FOR SEQ ID NO: 98:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8654 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

CGAGACAACA AGATGAAGAA AAATTTGCCC TATCGTTTGT GCGGCTTGCA AGTGTAGCAC	60
TTCTTGACGC CTGTGGAGAA GTGAAGTCTG GAGCAGTCAA CACTGCTGGT AACTCAGTAG	120
AGGAAAAGAC AATTAAAATC GGGTTTAACT TTGAAGAATC AGGTTCTTTA GCTGCATACG	180
GAACAGCTGA ACAAAAAGGT GCCCAATTGG CTGTTGATGA AATCAATGCC GCAGTGGTAT	240
CGATGGAAAA CAAATCGAAG TAGTCGATAA AGATAATAAG TCTGAAACAG CTGAGGCTGC	300
TTCAATTACA ACTAACCTTG TAACCCAATC TAAAGTATCA GCAGTCGTAG GACCTGCCAC	360
ATCTGGTGCG ACTGCAGCTG CGGTAGCGAA CGCTACAAAA GCAGGTGTTT CATTGATCTC	420
ACCAAGTGCG ACTCAAGATG GATTGACTAA AGGTCAAGAT TACCTCTTTA TTGGAACTTT	480
CCAAGATAGC TTCCAAGGAA AAATTATCTC AAAGTATGTT TCTGAAAAAT TAAATGCTAA	540
GAAAGTTGTT CTTTACACTG ACAATGCCAG TGAATATGCT AAAGGGATTG CAAAATCTTT	600
CCGCGAGTCA TACAAGGGTG AAATCGTTGC AGATGAAACT TTCGTAGCAG GTGACACAGA	660
CTTCCAAGCA GCCCTTACAA AAATGAAAGG GAAAGACTTT GATGCTATCG TTGTTCTCTG	720
TTACTATAAT GAGGCTGGTA AAATTGTAAA CCAAGCGCGT GGCATGGGAA TTGACAAACC	780
AATCGTTGGT GGTGATGGAT TCAACGGTGA GGAGTTTGTA CAACAAGCAA CTGCTGAAAA	840
AGCATCAAAC ATCTACTTTA TCTCAGGCTT CTCAACTACT GTAGAAGTTT CAGCTAAAGC	900
TAAAGCCTTC CTTGACGCTT ACCGTGCTAA GTACAATGAA GAGCCTTCAA CATTTCAGC	960
CTTGGCTTAT GATTGAGTTC ACCTTGATAG AAACGCAGCA AAAGGTGCTA AAAATTTCAGG	1020
TGAAATCAAG AATAACCTTG CTAAACAAA AGATTTTGAA GGTGTAAGTGT GTCAACAACG	1080
CTTCGATGCA GACCACAACA CAGTCAAAAC TGCTTACATG ATGACCATGA ACAATGGTAA	1140
AGTTGAAGCA GCAGAAGTTG TAAAACCATG ATAGAAAAAT GTTGAAATAG GGAATGAGCC	1200
TTTGACTCAC TCCCTGTTTC GATATTTAAT ACTCTTCGAA AATCTCTTCA AACTGCCTCA	1260

747

ACGTCGCCTT GGATTATATA TGTGACTGAC TTCGTCAGTC TTATCTACAA CCTCAAAGCA	1320
GTGCTTTGAG CAACCTGCGG CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATAA	1380
GAACCTATCA AAAAGTGAGG GAAAACCCCTC GGAATTATAA ATAGAAAGAG TGAATCTTAT	1440
GCTCCAACAA CTCGTAAATG GTTTGATTCT AGGTAGTGTT TACGCGCTGT TAGCCCTAGG	1500
ATATACCATG GTTTACGGAA TTATCAAGCT CATCAACTTC GCCCATGGTG ATATTTATAT	1560
GATGGGAGCC TTTATCGGTT APTTCTTGAT CAATTCTTTC CAAATGAATT TCTTTGTAGC	1620
GCTTATTGTA GCTATGCTAG CGACAGCTAT TCTTGGTGTC GTGATTGAGT TTCTTGCTTA	1680
CCGACCTTTG CGCCACTCTA CTCGTATTGC TGTTTTGATT ACGGCTATTG GGGTTTCTTT	1740
CCTATTGGAG TATGGAATGG TCTATCTGGT TGGTGCCAAT ACCCGTGCCT TCCCTCAAGC	1800
GATTCAAACA GTTCGATATG ATTTGGGACC AATTAGCTTA ACAAATGTGC AGTTAATGAT	1860
TTTGGCCATT TCCTTGATTT TGATGATTTT GTTACAAGTC ATTGTCCAAA AGACTAAGAT	1920
GGGGAAGCC ATGCGTGCAG TATCAGTAGA TAGCGACCGG GCGCAATTGA TGGGGATCAA	1980
TGTAAACCGT ACGATTAGCT TTACCTTCGC TTTGGGTCTT GCTCTTGCGG GTGCGGCTGG	2040
TGTTCTGATT GCTCTTTATT ATAACCTCTT TGAGCCTTTG ATGGGGGTTA CTCCAGGTCT	2100
TAAATCTTTC GTTGCCGCAG TACTTGGTGG TATCGGAATT ATTCTTGGTG CGGCTCTTGG	2160
TGGCTTTGTG ATTGGTCTAT TGGAAACCTT TCGCACTGCC TTTGGGATGT CAGATTTCCG	2220
TGATGCCATT GTTTATGGAA TCTTGTGTTT GATCTTGATT GTCCGCCCAG CTGGTATCCT	2280
TGGTAAGAAT GTGAAAGAGA AGGTGTAAAC GATGAAGGAA AATTTAAAAG TTAATATTCT	2340
ATGGTTACTC CTTTTGTTAG CTGGCTATAG CTTGATTAGT GTACTGCTTT CAGTCGGAGT	2400
ACTTAATCTA TTCTATGTAC AGATTTTACA ACAAATTGGA ATTAATATTA TTTTGGCTGT	2460
TGGTCTCAAC TTAATCGTTG GTTTTTCAGG ACAATTTTCA CTGGTCATG CTGGTTTCAT	2520
GGCGATTGGT GCCTATGCAG CAGCTATTAT TGGTTCTAAA TCACCAACCT ACGGTGCCTT	2580
CTTTGGAGCT ATGCTGTAG GGGCTTTGCT TTCAGGAGCA GTTGCCTTAC TTGTCGGCAT	2640
TCCAACCTTG CGCTTGAAGG GGGACTATCT TCGGTAGCA ACTCTGGGTG TTTCTGAAAT	2700
TATCCGTATC TTTATCATCA ATGGTGGAAG CCTTACAAAT GGTGCGGCAG GTATCTTAGG	2760
GATTCTTAAC TTTACAACCT GGCAATGGT TTACTTCTTT GTCGTGATTA CAACCATTGC	2820
AACCTTGAAC TTCTTGCGTA GCCCAATTGG TCGTTCAACC CTCTCTGTTT GTGAAGATGA	2880
AATCGCTGCT GAGTCAGTTG GGGTTAATAC GACTAAAATT AAAATCATCG CTTTTGTCTT	2940
TGGTGCCATT ACTGCAAGTA TTGCTGGGTC ACTTCAGGCA GGATTTATCG GGTCTGTTGT	3000

748

ACCGAAAGAT TACACCTTCA TCAACTCAAT CAACGTTTGG ATTATTGTTG TATTTGGTGG	3060
ACTCGGTTCC ATTACAGGTG CGATTGTTTC GGCTATTGTT CTGGGAATTT TGAATATGCT	3120
TCTCCAAGAT GTTGCTAGTG TGGTATGAT TATTTACGCT TTGGCCTTGG TATTGGTAAT	3180
GATTTTCAGA CCAGGTGGAC TCCTTGAAC ATGGGAACTG AGCCTATCAC GTTCTTTAA	3240
AAAACTAAG AAGGAGGAAC AAAACTAATG GCATTACTTG AAGTAAACA GTTAACCAA	3300
CATTTTGGTG GTCTAACAGC TGTGGAGAT GTGACTCTTG AATTGAACGA AGGGGAACTG	3360
GTTGGATTAA TCGGTCCAAA CGGAGCTGGG AAAACCACCC TTTTCAACCT TTTGACCGGT	3420
GTTTATGAAC CAAGCGAGGG AACAGTAACC CTAGATGGTC ACCTTTTGAA TGGGAAATCA	3480
CCTTATAAGA TTGCCTCTTT GGGACTTGA CGTACTTTCC AAAATATCCG TCTCTTTAA	3540
GATTTAACAG TTTTAGATAA TGTTTTGATT GCTTTTGGA ACCATCACAA ACAGCATGTT	3600
TTTACTAGTT TCTTACGCTT ACCAGCTTTT TACAAGAGTG AAAAAGAATT AAAGGCTAAA	3660
GCTTTGGAAT TGTGAAAAT CTTTGATTTA GATGGTGATG CAGAGACTCT TGCTAAAAAT	3720
CTTCTCTACG GACAACAACG TCGTTTGGA ATTGTTCTGT CCCTTGCTAC GGAACCTAAA	3780
ATTCTCTTCT TAGATGAACC AGCAGCAGGT ATGAACCCAC AGGAAACAGC CGAATTGACT	3840
GAGTTAATTC GTCGTATCAA AGATGAGTTT AAGATTACAA TCATGTTGAT TGAACACGAT	3900
ATGAATCTGG TCATGGAAGT AACAGAACGT ATCTACGTAC TTGAATATGG CCGTTTAATC	3960
GCTCAAGGAA CTCCAGACGA AATTAAGACC AATAACGCG TTATCGAAGC TTATCTAGGA	4020
GGTGAAGCCT AATGTCTATG TTAAGAGTTG AAAATCTTTC TGTGCATTAC GGTATGATCC	4080
AAGCAGTTCG TGATGTAAGC TTTGAAGTTA ATGAAGGAGA AGTTGTTTCC CTTATCGGTG	4140
CCAACGGTGC AGGTAAAGCA ACTATTCTTC GCACCTTGTG AGGTTTGTTT CGACCAAGTT	4200
CAGGAAAGAT TGAATTTTGA GGTCAAGAAA TCCAAAAAAT GCCAGCTCAG AAAATCGTGG	4260
CAAGTGGTCT TTCACAAGTT CCAGAAGGAC GCCACGTCTT TCCTGGCTTG ACTGTTATGG	4320
AAAACTTGA AATGGGAGCT TTCTTAAAGA AAAATCGTGA AGAAAATCAA GCTAACTTGA	4380
AGAAGGTTTT CTCACGCTTT CCTCGTCTTG AAGAACGGAA GAACCAAGAT GCAGCCACTC	4440
TTTCAGGGGG GGAACAACAA ATGCTTGCCA TGGGACGCGC CCTCATGTCA ACACCAAAAC	4500
TTCTTCTTTT AGATGAACCA TCAATGGGAC TTGCCCCAAT CTTTATCCAA GAAATTTTTG	4560
ATATCATTCA AGATATTCAG AAGCAAGGAA CAACGGTCCT CTTGATTGAA CAAAATGCCA	4620
ATAAAGCACT TGCAATCTCT GACCGAGGAT ATGTACTGGA AACAGGGAGA ATCGTCCTAT	4680
CAGGAACAGG AAAAGAACTC GCTTCATCAG AAGAAGTCAG AAAAGCATAT CTAGGTGGCT	4740
AAAACAATCC AGTGGATTGT TTTAGTCGGC AGATGGAGAT TACGAAGTAA TCATCAATAT	4800

749

AGTCCGGGGG ACCTTTTTAG TCGGTAGATT GAGATTGCAA ACAAATCTGC ATCTACATTG	4860
AAAGCTTAAT TTCTAATAAT TGAAAAAATC GAATGAAAAA TTTCTTACCT TCATTACAG	4920
AGCTCGATTT CAGAGCTCTT TTTGCTAGCT TATTCACTACT TTTCTGAATT TCGAAAAAGA	4980
AATGTAAGCG TTTGATAGAT TTACAAAAAG ATTGTATAAT AGGGATAAGA ATAGAAAAGG	5040
AGAAGTCTCA TGGCACTTAA AGATTTTATG ACCCGCAAGG TAGTTTATAT TAGTCCAGAT	5100
ATAACAGTAT CTCATGCAGC AGATTTGATG AGAGAGCAAG GTTTGCACCG TCTGCCTGTT	5160
ATCGAAAATG ATCAATTAGT TGGTTTGGTG ACTGAGGGAA CCATTGCACA AGCAAGTCCA	5220
TCTAAAGCAA CAAGTCTTTC TATCTATGAG ATGAATTATC TTCTGAATAA GACAAAAGTA	5280
AAAGATGTCA TGATTGCGGA TGTGTGCTACT GTCTCAGGCT ATGCTAGTCT AGAAGATGCA	5340
ACTTATCTGA TGTGAAAAA TAAGATTAGT ATTCTCCCTG TCGTAGATAA CCATCAAGTA	5400
TACGGAGTTA TTAAGTACCG TGACGTTTTT CAAGCCTTTC TTGAAATTGC AGGTTATGCG	5460
GAAGAAGGGA TTCGTGTACG CTTTGTTACA GAAGATGAAG TTGGTGTCTT TGGAAAAATT	5520
GTTCCTTTGA TTGTAGAAGA AAATTGGAAT ATCTCCCATC CAGTCAATAT TCCGCGTAAG	5580
GATGGTAAGG TGATTATCGA AGTGCAAATC GATGGATCAA TTGATTTACC AGCCTTGAAA	5640
GAAAAATTG AAGCAAATGG TATTCAAGTG GAAGAAATCG CTCGCACTTC AGCAAAAGTC	5700
TTGTAAGAAG GGAAGCCCAA AGGCTTCTTT TTTCATGAAA AGGGGATTAG AGCAAAAGAT	5760
GGAAAGAAAT GATAAAATAT GCTATAATGA AATAATGTAA AAAAGGAGTA TTTATGGACA	5820
TTTCAGTAAT TCGTCAGAAA ATTGACGCAA ATCGTGAAAA ATTAGCTTCT TTCAGGGGGT	5880
CTCTTTGACC TCGAAGGGCT AGAGGAAGAG ATTGCCATCT TGGAAAAACA GATGACAGAA	5940
CCTGATTTTT GGAACGATAA TATTGCGGCC CAAAAACGT CGCAAGAATT AAATGAATTA	6000
AAAAACACTT ACAATACCTT CCATAAGATG GAAGAGTTGC AGGATGAAGT CGAAATTTTA	6060
TTGGATTTTT TGGCTGAAGA CGAGTCAGTG CATGATGAAC TGGTAGCGCA GTTAGCCGAA	6120
CTTGATAAGA TAATGACCAG CTACGAGATG ACTCTACTCT TGTCAGAACC TTATGACCAC	6180
AACAATGCCA TCTTGAAAT CCATCCAGGT TCTGGTGGTA CTGAGGCGCA GGAAGGGGT	6240
GATATGTTGC TTCGTATGTA TACTCGTTAT GGTAAATGCTA AAGGCTTTAA AGTGAAGTG	6300
TTGGATTACC AAGCAGGTGA TGAGGCTGGT ATTAAGTCGG TAACTTTATC ATTTGAAGGG	6360
CCTAATGCCT ATGGTCTCCT CAAGTCAGAA ATGGGTGTTT ACCGCTTAGT GCGAATCTCA	6420
CCATTTGACT CTGCCAAACG TCGCCATACC TCTTTCACAT CTGTAGAAGT GATGCCAGAA	6480
TTGGATGATA CTATTGAAGT GGAAATCCGT GAAGATGATA TCAAGATGGA TACCTTCCGT	6540

750

TCAGGTGGTG CCGGTGGACA AAACGTCAAT AAGGTTTCAA CAGGTGTACG TTAAACCCAC	6600
ATTCCAAC TG GAATTGTTGT CCAATCAACA GTAGATCGTA CCCAGTATGG AAATAGAGAT	6660
CGTGCCATGA AGATGTTGCA GGCTAAGCTC TATCAAATGG AGCAAGATAA GAAGGCTGCG	6720
GAGGTAGATT CTCTCAAAGG TGAGAAAAAG GAGATCACTT GGGGAAGCCA AATCCGTTCT	6780
TATGTCTTCA CGCCTTATAC TATGGTAAAA GATCACCGAA CTAGCTTTGA GGTGCTCAG	6840
GTAGATAAGG TTATGGATGG GGACCTAGAT GGTTTTATCG ATGCTTATCT CAAGTGGCGA	6900
ATTAGCTAAG ATAGAAAGGA ACTCACATGT CAATTATTGA AATGAGAGAT GTCGTTAAAA	6960
AATACGACAA CGGAACAACT GCTCTACGCG GTGTTTCGGT TAGCGTTCAA CCGGGGGAAT	7020
TTGCTTACAT CGTAGGACCT TCAGGAGCAG GGAAGTCAAC TTTTATTCGT TCTCTGTATC	7080
GTGAAGTAAA AATCGATAAA GGAAGCCTAT CAGTTGCTGG TTTTAATCTG GTTAAGATCA	7140
AAAAGAAAGA TGTCCCGCTT CTACGTCGTA GTGTTGGGGT TGTCTTCCAG GATTATAAAT	7200
TGTTACCAAA GAAACTGTC TATGAAAATA TTGCTTACGC TATGGAAGTA ATCGGGGAAA	7260
ATCGCCGTAA TATCAAAAGA CGAGTGATGG AAGTTTTGGA CTGCGTTGGA TTGAAGCATA	7320
AGGTTTCGTT TTTCCCAAAT GAACTCTCAG GTGGGGAGCA ACAGCGGATT GCGATTGCGC	7380
GTGCAATTGT AAATAATCCC AAAGTATTGA TAGCTGATGA GCCAACAGGA AATCTGGATC	7440
CGGATAATTC ATGGGAAATT ATGAATCTCT TGAACGGAT TAACyTACAA GGAACAATA	7500
TTTTGATGGC GACTCATAAT AGCCAGATTG TAAATACCTT GCGCCACCGT GTCATTGCCA	7560
TTGAAAATGG CCGTGTCTT CGTGACGAAT CAAAAGGAGA GTATGGATAC GATGATTAGT	7620
AGATTTTTTC GCCATTTATT TGAAGCCTTA AAAAGTTTGA AACGAAATGG TTGGATGACA	7680
GTAGCTGCTG TCAGTTCAGT CATGATTACT TTGACCTTGG TGGCAATATT TGCATCTGTT	7740
ATTTTCAATA CAGCGAACT AGCTACAGAT ATTGAAAATA ATGTCCGTGT AGTAGTTTAT	7800
ATCCGAAAGG ATGTGGAAGA TAATAGTCAG ACAATTGAAA AAGAAGGTCA AACTGTTACA	7860
AATAATGACT ACCACAAGGT ATATGATTCT TTGAAGAACA TGTCTACGGT TAAAAGTGTT	7920
ACCTTTTCAA GTAAAGAAGA ACAATATGAA AAATTAACCG AGATAATGGG AGATAACTGG	7980
AAAATCTTTG AAGGAGATGC CAATCCTCTC TATGATGCCT ATATTGTAGA GGCAAACACT	8040
CCAAATGATG TAAAACTAT AGCCGAAGAT GCTAAAAAAA TTGAAGGTGT CTCTGAGGTT	8100
CAAGATGGCG GTGCCAATAC AGAAAGACTC TTCAAGTTAG CTTCAATTTAT CCGTGTTTGG	8160
GGACTAGGGA TTGCTGCTTT GTTAATTTTT ATCGCAGTTT TCTTGATTTC AAATACCATT	8220
CGTATTACCA TTATTTCCCG CAGTCGCGAA ATTCAAATCA TGCGCTTGGT CGGAGCTAAA	8280
AACAGTTATA TCCGTGGACC GTTCTTGTTA GAAGGAGCCT TTATCGGTTT ATTGGGAGCT	8340

751

ATCGCACCAT CTGTTTGGT CTTTATTGTT TATCAAATTG TTTACCAATC TGCAACAAA	8400
TCGTTGGTAG GGCAAAATCT ATCCATGATT AGTCCAGATT TATTTAGTCC GTTGATGATT	8460
GCCCTACTAT TTGTGATTGG GGTTTTCATT GGTTCATTGG GATCAGGAAT ATCCATGCGC	8520
CGATTCTTGA AGATTTAGGT AAAATAGCTG CTTTATGAG GAGATTGTAA AATCTCCTTT	8580
TTTGCTACAA GAGTTTGA AAAGAGATGC GCAGAAGAAA AGAGCTTCCA AAGAAGTCCC	8640
CCAGAGAAGA CTTC	8654

(2) INFORMATION FOR SEQ ID NO: 99:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 19718 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

TGTCGCGTCA AAATCATTAC TATGGCTATG TATAGCCCTT ACTATGACTT GGCTAAACAC	60
GTTTCGCTTTC AAATTTCTAG GCTCAGGCTG AAACAGTCTC CCAGGCTGTT CACTCCCGAA	120
TGCTAAAATC GTTCTTGATC GCTTTCACAT TGTACAACAT CTTAGCCGTG CTATGAGTCC	180
TGTGCATGTC CAAATCATGA ATCAGTTTCA TCGAAAATCC CATGAATACA AGGCTATCAA	240
GCGCTACTGG AAATCATTAC AACAGGATAG CCGTAAACTG AGTGATAAGC GATTTTATCG	300
CCCTACTTTT CGCATGCACT TAACAAATAA AGAAATTCTT GACAAGATTT TAAGCTATTC	360
AGAAGACTTG AAACACCACT ATCAGATCTA TCAACTCTTA CTTTTTCACT TTCAGAACAA	420
AGACCCTGAG AAATTTTTCG GACTCAATTGA GGACAATCTG AAGCAGGTTT ATCCTCTTTT	480
TCAGACTGTC TTTAAACCT TTCTCAAAGA TAAAGAAAAG ATTATCAACG CCCTTCAACT	540
ACACTATTCT AATGCCAAAC TGGAAGCGAC CAATAATCTC ATCAAACCTA TCAAGCGCAA	600
TGCCTTTGGT TTTGAAACT TTGAAAACCT CAAAAACGG ATTTTATCG CTTTGAACAT	660
CAAAAAAGAA AGGACGAAAT TTGTCCTTC TCGAGCTTAG CTGACTTCAA CCCACTACAG	720
TTGACAAAGA GCCTAATTC CATAAAATT GACATGGAAA TTATAAAACC ATTACTAGTT	780
TAGTCCTTTT TGATAACGTG CCAATTCGGC TTGGTTCGCC CAAACATAGT GACCTGGACG	840
GATTTCTACC ATAGATGGCT TATCAGTCTC ATAGTCGTGT TGACTTGGAT CGTAAACCTT	900
CAAGACCTTC TTACGTTCCA AGATTGGATC TGGGATTGGT ACCGCTGAAA GCAAGGCTTG	960
AGTATATGGG TGAATTGGAT TGTAAACAA TTCTTCTGTT TCTGCAACCT CTACAATAAC	1020

752

ACCCTTGTA	ATAACTGCCA	TACGATCTGA	AATAAAGCGA	ACAACCGACA	AGTCATGGGC	1080
GATGAAGAGA	TAGGTCAGGC	CGAGCTCTTT	TTGGAATTTT	TTGAGCAAGT	TCAAGACTTG	1140
GGCACGTACA	GAAACGTCCA	AGGCTGAAAT	TGGCTCATCT	GCAATAACAA	AGTCTGGTTG	1200
CATGACCAAG	GCACGGGCAA	TACCGATACG	TTGACGTTGA	CCGCCTGAGA	ATTCATGAGG	1260
GTAACGAGTC	AAGTGCTCAG	CAAGAAGACC	TACTTCACGG	ATAATATTTT	GAACCTTCTC	1320
TTTACGTTCT	TCTTCATCCT	TAAATAAAGC	GTGATTGTAA	AGACCTTCAG	AAATAATATA	1380
ATCAACAGTC	GCACGTTTAT	TCAAACCTGC	GGCAGGGTCT	TGGAAAATCA	TCTGGATTCTG	1440
ACGAATCAAT	TCCGCAGCTT	GTTCACGCGA	TTTCTTACCA	TTAATCTTTT	GACCATCAAA	1500
AATGATATCT	CCATTACTTG	TATCATTTAG	ACCGATGATA	GCACGACCAA	TAGTTGTTTT	1560
CCCCTACCG	GACTCACCTA	CAAGCGAGAA	AGTTTCTCCC	TTGTTGATAA	AGAAGTTAGC	1620
ATTTTAAACC	GCGACAAACT	TCTTACTTCC	TTCACCGAAG	GAAATTTCTA	AATCTTTGAT	1680
TTCTACTAAT	TTTTAGACAA	TTTCTTCTCT	CCTAGTCAGC	CAGATGGGCA	AATCCCATT	1740
TTTCACGGAT	CTTATCATGG	AGATTGCAA	TCACAGCTGG	TTTTTCTACT	TTCGGAGCAT	1800
CCTCATGAAG	AAGCCAAGTT	TTAGCCCAAT	GTGTCTCTGA	TACTGAGAAT	TGAGGAGCTT	1860
TTTGTTGAA	GTCAATCTGC	ATTGCGTAGT	CAGAACGCAA	GGCAAAAGCA	TCCCCTTTCA	1920
GGTCAGTATA	AAGTGACGGA	GGTGTTCCTG	GGATTGAGTA	AAGATCCCCT	TTATCATCAG	1980
CAAGCTGAGG	CAAGCTAGAC	AAGAGACTCC	ATGTATATGG	ATGGCGAGGG	TCATAGAAGA	2040
CTTCTCAAC	CGTTCCATAC	TCAACGATTT	CTCCTGCATA	CATAACCGCT	ACCTTATCCG	2100
CAATACTTGC	CACCACACCA	AGGTCGTGGG	TAATAAAGAT	TGTTGTGAAA	TGATACTCGT	2160
TTTGTAAGA	TTTTAGCAAA	TCAATAATCT	GAGCTTGAAT	AGTTACATCC	AAGGCAGTTG	2220
TTGGCTCATC	ACAGATCAAG	ACATCAGGTC	GGCAGGCAAG	GGCAATAGCA	ATAACGATAC	2280
GTTGACGCAT	TCCTCCAGAA	TATTGGAATG	GGTATTCAAT	AAAACGTCTA	TCTGCGTCTG	2340
GAATGCCAAC	CTTATTATG	TAGTCAATGG	CCAATTCTTT	CGCTTCTTTA	GCTGTTTTTC	2400
CTTGGTGT	TACAATAACT	TCTGTAATCT	GACTACCAAT	TGTTTTAATG	GGGTCCAAAC	2460
TAGTCATTGG	GTCTGGAAG	ATAGTCGCAA	TCTTAGCACC	ACGAATTTGT	TCCCAATCCT	2520
TGTGAGAAGA	TAAAGCTGTC	AAGTCCTGAC	CACGGTAGTC	AATACTACCT	TGGGCAATAC	2580
GACCATTTTC	TTGAGCATA	CCTGTGAAGG	TCTTTGTCAA	AACAGATTTA	CCTGATCCTG	2640
ACTCACCTAC	CAAGGCTAAT	ACTTCTCCTT	CGACTAGTTC	AAGGGAAACG	CCGCGAATGG	2700
CTGTCAATAC	TTTGTACGA	ACGTCAAATT	CCACGACAAT	ATCGCGAGCA	GTCAAAATTA	2760
CATTTTTTTC	TTTTGTCAAT	TCTACTCCTA	TCTATGTGTA	CGTGGATCAC	TAGCATCCGC	2820



753

TAAGTTTGA CCAACTACGA AAAGGGACAA GGATACCAAG ACAAGGGTTG TCAATGGAAT	2880
CCAGAACAAG TAAGCATTGG TTGTTACGTT TTGTGAATAA TCCGAAATCA AACGACCCAA	2940
ACTTGGCACT GTAATCGGTA ATCCAAGACC GAAGAAAGAC AAGAAGGCTT CGTATGAGAT	3000
AAAGCTTGA AGCATTTGAG TCATGGTTGT CACAATAACA GATACCAATT GAGGCATGAT	3060
ATTTTGGCA ACAATCTTCA AGGTTGGTGT TCCCAAAGTA CGTGACGCCA AGTTGTATTC	3120
CAAGTCACGA TAGCGCAAGA TTTGCACACG GATCATGAAG GCAATACCAA TCCATGTTGT	3180
TACGCTCATG GCAAAAATCA GATTCCAGAA TCCAGCTCCG ATTGAGTAAG TCAAGACAA	3240
AACAATCAAA AGAGGTGGGA TGTGAGAT GACGTTGTAA ACTTCCATCA TGACACGGTC	3300
AACTGATTTT GAAATACCCC AAATACCACC GACAAAAACA CCGATAACCA AGTTAATCAC	3360
TGTCGCAATC ACAGAAATGA GGATGGAGTT ACGAGCTCCG AACCAGACAC CGTCAAAGAG	3420
CGATTTACCG TTACTGTCAG TACCGAACCA ATGCTCCGCA TTTGGCTTGA TATAACGAAC	3480
ACTAAAGTCG TTTACCTTGC TGACATCATT GAAATCAAAC TTAGAAAACA TTGGGTAGAT	3540
GAACTTATC AAAATGATGG CTACCAAGAT TCCCAACATG ACTACAGTTG ATTTTCTT	3600
CATAAATTGT TTAACACTG ATTTCCAGTA AGAATATGCT GGCGCATCAA TAGTTTCAGA	3660
GGCAAAATCG TCACGTTTAA CAACTGAAA TTTTCTTTA TCGATTGTAG ACATTATTG	3720
CCTCCTTCT CAGTCAATTT AATACGTGGG TCAATAATAG TCATCCAAAT ATCTCCCAA	3780
AGCGTGAGA AGATAGAAAT ACATGTAAAG ATGAAGACAA GACCAACGAC CATAGAGTTA	3840
TTAGATGCTT TTACAGAGTC AATCAACATT TTACCCATAC CTGGGAAGGC GAAGACTGTT	3900
TCAGTAAGGG TTGCACCACC GATAACCCCA ATAATGGCAG CAGGAATTCC TGAAACCAGC	3960
GGAACCATGG CATTTTAAA GATGTGTTG TTTGAAATTT CTTTTTCAGA CAAACCTTTT	4020
GCAGGAGCGA AACGAACAAA GTCTTGAGAT TGCAAGTCAA TCATGTAACG ACGAATCCAA	4080
ATGGCTGTAC CAGGAGCACC CAACAAACCA AGGATGACTG CTGGTAAAC GTAAGAACGC	4140
CAATCTCCAG CTCCAAGAT AGGGAATGAA TCTGGAAGGG CAATAGATGA TCCAATCAAT	4200
CGAACGATGT AAACCAAGGC AATCGTTGGA AGAGCAAGCA AGAAGGTCAA AGCCCCGTGT	4260
GAGAGGCTAT CAATCCAAGT GTTCTTGAAA CGAGCCATGG CTGAACCAAG TGGCACGGCA	4320
AGAGCATAGG CAAGAACCAA ACCAATCAAA CCAGTAATAG CAGAGCTGAC AATCATAGAT	4380
GGATATTGGT AATTACTTTC AGTCGCTGTA TAAGGATCAT CTTTCCATA GCTAGCTACT	4440
TCACGAGACT CAGCCTGACT AGGTGACTTG TAGGTTCTTG AGTAAATATT TACAGAAGAC	4500
GTTTCTTAC CTGTTGGGAA CTGAACTTGG GCAGTTTGG TTTGTCCTTG ACCTTGAGTA	4560

754

ATAACCTGAA	GAACTGGTGT	ATTAGCATAG	GTTGGGTAAG	AGTCACCTAA	ATTCAAGTTC	4620
ACAAAGTTTT	GATGAACAAA	TGGGAACTGA	CTGTTAAAGT	ACAAGAGATA	TTTATGTTTA	4680
GTTCTGAAC	CGACCAATGA	CCATCCGATA	GCTGGATCAT	TTTCAAAACG	AAGGTAGCGT	4740
TTCAAGTCTG	GATTTTCAGG	GTCTTGATT	TTATTTGTAT	GGTCAATGTC	AATCAAGTTA	4800
GCATAGAAGT	GAAAAACACG	TTCAAAAATT	GGAATTTTAC	GAGTAGCATA	GAATTGACCA	4860
CTTTTCAGTAA	ATTCTCCCAA	AGTCCAACCA	TGACCTAATT	GATTGATGTA	CTTTTCATAA	4920
ATAGCTTTAT	TGGTCGCATT	TGCTTCTACT	GTTACAGAAG	AATCCATGCT	ACTTGCCTTT	4980
TCTTGCAACT	CTTTAGTATC	GTAATACTCA	ATGTAGCCCA	TACGCTCAA	CACAGTATTT	5040
TCATAGTTAT	CACGTTTATC	AGCCGTTGTC	GCAATTTTAT	TATAGTTAGG	ATCCTGCTTG	5100
AAAATCAATT	TTCGAGGAAC	CAAGGTATAG	ATAATCGTGT	AGGTCAAAGT	CGTTACTAAG	5160
AAAATCGAAA	CCAATGACCG	CAAAACACGC	ATAAAAATAT	ATTTTTCAT	ATTATTTCTT	5220
TTAAAAATCC	CAAAAGAACC	TTCTCCTCAT	GGAGAGAAAG	TTCTATTAGA	AATTATTTAC	5280
TTACATGAC	TTGCCAATTC	TTTTTGAGCT	TTCTCATTTG	ATTCAGCTTT	TTCTTTCAAC	5340
CATTTTTCAC	GAGCTTTTTT	ATACTCTTCC	TTAGTCACCA	CTTTATCTTG	TGATTTCAAA	5400
TATTTGAAGT	AAACATCTGA	CCCCTTAGAG	CCTGTTTGCG	CAGAAGCTCC	AGTAAATGGA	5460
ACAATTCGTG	AAAGCACTGG	TGCTGCACCA	GAAGAAGCCA	TAGCAGGAAT	AAAGAGTGAA	5520
CTATCTGTCA	ACCATGCTTG	AGCCGCTGCA	TATTTTTTCAT	AACGGACATT	CAAGTCGCTT	5580
GTCTCTCTGG	CAGCTTCATC	AACTAATTTA	TCGTATTCTT	TCAAACCAAC	TTGAACTACT	5640
GAAGGGCTAT	TTGGATTATC	AAATCCTAAA	TATGTTTTTG	TAGTTTCACT	GCTAGTTGTT	5700
TTTAAATAT	CCAGGTAAGT	AGATGGGTCT	TGATAGTCTG	GCCCCCATGA	AACTCCTCCT	5760
GATACATCCC	AATCCTCAGA	TGAAGCATTG	GCAGCATAGT	AAGTAATATT	AAGGAATTCA	5820
TCACTTGTC	TTTGTTGAAT	ATCAACAACG	ACATTTTCAA	CACCAAGAAC	TGTTTCTACA	5880
GATTGTTTAA	AGGACTGAAT	ACGAGATATG	TAGTTTTTTG	ATGCTTGGTC	TACTGGAACG	5940
TCCAGATGAA	TAGGAAACTG	AACGCCGTCT	GCTTCTAAAG	CTTTCTTAGC	TTTCGCAAAC	6000
TCTGCCTTGG	CCTTGTGAGC	ATTGAATAAA	CCATCCTGCC	CATCAGCTAA	ATTCACACCT	6060
TTCCACTCAT	CACCATAAGC	AGGAAGTTGA	GCAGCGACTA	AATCACCAAA	GGTCTTCTCA	6120
CCAGCTGAAA	CAAAGTCTGG	TTTTACAAAT	AAATTACGAA	CTGCTAAAGC	TGCTCCATCT	6180
TTACCATTGA	TTTGAGCTGA	GTAAGCTGAG	CGATCAAGAG	CAAAATTCAA	GGCTTGACGG	6240
AAATCTTTGT	TAAGCAATGC	CTTCTTAGTA	GCTACTTTCT	CTGAATCTGT	AGTTTTAGAA	6300
GTATAGTTGT	AACTTTGGCG	ATCAATATTC	ACACCCAGAC	CAGCAATCCC	AGAGCCTGAT	6360

755

TGTGTGTAAT AGATATTGTC CTTGTATTCT TCTGCAACCT TAGAATAGTT GGAGCTGGTA	6420
GGGTAAAGAC GGGCATAACT ATAAGCTCCA CTAGTGAAGT TACGCTCTAG CGACTCCTGA	6480
TCTGATCCAT CATAGTAAGC TAGATTGATA GTATCTAGGT GGACATTTTC TTTATCCCAA	6540
TATTGCTCAT TTTTACAAA CTCTACAGAA GATTTTGCAG TCAACCCTTT CAACAAGAAT	6600
GGACCATTAT AAAGCAAGGA TGTCGGATCT GTTGGTTTAG CAAAATCGCT TCCTTTTGAT	6660
GTTTCGAATT-CTTCATTGAG AGGCCAGAAA ATAGAATAGG TCAACTTAGA GTTCAGAAC	6720
GGTTCAGGCT GGTTCAAAGT GTATTGTAAC GTATAATCAT CAACCGCCTT GACACCAACT	6780
GTTGAAAAAT CTGTTGAAGT TCCTGATAGA TAATCTGCCA AGCCTTTAAC CGAATTTTCA	6840
GCTAAATACA TAGCTTCTGA TTTTATCT GCTGCGTGT TTAACCGTT CACGAAATCT	6900
TTAGCCGTCA CCTCTGCATA TTCTTCTCCA TCAGAGGTAA ACCATTTAAC CCCTTTACGA	6960
ATCTTATAAG TGTAGGTCAA ACCATCCTTA GAGACTTCCC AATCCTCTGC AATGCGAGGA	7020
GCAAGATTAC CGTAATTATC GTTAGTGAAT AAACCATCAA TCCCATTGA AGTCACTACT	7080
GTTGTACTAT TTTTACTTGA AATCAGGTAG TCCAAGGTTT CTGGGTCTGC TGTATAAACA	7140
TAGCCATAAG CTTTAGGGGC TGATGAATCA GATGATTTTG AAGAACTGCA TGCTGCAAGT	7200
ACACCTGCTG CTAATAAAAC AAGACCTGCT GTAGCAAATA CACGATTTTT TTTCAATTTTC	7260
TACTCCTCTG TTTATGTGAA TTATAGATTG ACAACCATTA TATCACATTA TCCATTAAAA	7320
ATCAAAACAAA TTTTCAGAAT ATTTAGGCTT GTTGGCAGAA ATTTTTCATT TTTTGTGAAT	7380
ATATGATTCA AATTGTCGTT CGAAGTGTC AAGACTACAG TGAAAATAGG AAATTTGACG	7440
CAGAACTTTT GGAGTTTAGG AAGACATACA GTAAAATGAA ATACGGACGG AACAATGTGA	7500
TTTTGGAATT CAAATTAAAT TATAACAATA TTGTAGAAGT ATCATTCTAG TATTCAAGAT	7560
TCAGTTTACT ATGTCTTTTC ACACCAACCT TATCCCGAAT TCAATTACTT TTGTGATTTA	7620
CATATATAGA TTAAGACTAT CTTTATACT TTAATAATTC TCGCTACCTT ATCCACTATA	7680
TGCTCCTCGC TATCACGTTT CTATTCATAG CCTACGATTT CACTATTGCT TTCTCTGACA	7740
ATTCTTATTT CCTGCGTCAG ACTTAAACG ATCTATCCCC AGACCATTTT AATCCGCTAC	7800
CTCACGATAG TCAGGCTTGG GGAGCGCTAT TGTATTCACC GGTAGTGGAG CCCTACAGAG	7860
GACTTACACC TCAGATGCAC GACATGCCCA TCGTAAAAA AATCTCCTAC CCAAGGTAGA	7920
AGATTTCAAA CTTATAAAAC TTAATCCGTC ATGTCCGATA CCAACATTCG ATGCTCCAAT	7980
GGAATACTGC ACATACTAG CAAGAAAATA AAGCCTGACT GAATCCAGAA GAGAGCCAAG	8040
TCAAAAATTC CGTGCACAGC AACCCTGTA AGGAAAGATA GATAAAGGCC GATAATCGGA	8100

756

CGTTTCCCCG ACTCCTGACT CATATCCATC ATCAAGCGAA CAGGAGCAAC AGAAGACAAA	8160
ACTAATAAAA TAGTCCCCAC AATTCCGTAA CTCAGAATCG TATCAATATA AAGACTGTGG	8220
GCATGTTTCAT GATAAGGAGC ATGTATCCGA GGATAAGAGT TCATATAGGT CAATGGCCCT	8280
TCACCCCAAA AAGGATTTTG CTAAACAAG GCCATCCCAG CATCCCAGAT AGAAATGCGT	8340
TCTTCCATAG AAGAGTCTAA AGTACCCATT CGAACTCCCA AATCACTAGA AAAGAGGAAA	8400
CTCAAACCAA TCGCGAAGAC CCAATACTA AGCCAAAAGG CCTTCCAGTT TTTAATAGTC	8460
GTAAGAGAT AGATAATTGC TCCAGCGATA ATAGCAGGAA AGGCAGTTCG ATTTTGAGTA	8520
AAGTTCAAAC CAAAGAGATT AACAAAGCCT GCAATCACAC AGAATACTTT CAACCAATTC	8580
AACTTGGTCG TTGTAAACAG ATAGAAAGCA ATCATAATAC AGAAACAACA AATAATTCCA	8640
TAATAATTAG GATTAAAGAA GGTCACTTCT GCCCGGTTCT GATGCCACAC CTGCATATTG	8700
GGTGAAGAA AAGCATAGTT AAATTTCTTC ACAATTTGGA AATGTTCTAA ACTGGCAAAA	8760
GCAGCTGACA AGACACTACC AAACAAGACA AACTGCAAAA TCAATCGAAA GAATTTATGG	8820
GATAAAATCG ACTGATAGTG CAAAAGAAA ATAGTAAATA GAAACATTCC TACTGAAGCC	8880
ACAAGACCCA TCCAATTTTG TGCAAGAATG GATATAACAG TACTATAGCT AAGAAAAAGA	8940
AGCAGCATCG GATGCTCCCC CATTTTCTGA AGAATACTTT TCATGTCTCC TGTA AAAATC	9000
AAACTGATAA TATATAACA GAGTACAACT AAAAAAGAT AAAAGGGTAA AAAGATACTC	9060
AGGATAATTC CCAATAAAAT CAGCTCTTTA CTAGACAACC CCTTCAGCTT TTCAATAAAG	9120
CCTATTGATT TCAAAATGAA TCCTTTCTCT CCAATCAGC TGATTCAGAT AATAGTAAGC	9180
TATCCTATAT TGTACCACTT TTTAGCAAT TTGAAAACAA AGGAAACGTT TTCCAAAATA	9240
AAAACCTAT TTTATCCACC ATATCAAGGC TTCAAATGA TACTTCAACT CCATTCTCAA	9300
TTACCCGATA AGTCTGATTT TGCAAATCAA TTTCTACTAC TGCTGTTACG GACTTATCTT	9360
TATTTTGACG TTTGATTACA ATGCTGTGAG CTGTTGGTGT CTCTATCTCA GTAGTCCCTT	9420
CTAGATCAAA GGCTTCTGAA CGGTTACGGA AAGAAAATAG ATTGAGAAGG GCCTTCACAA	9480
CAGGTCGTTG CACTTCTTTT GCTATTTTCT CGTTGCTATA GTAATGACGA TTAATATTTT	9540
GACCTTCTTT AGTTTCTTCT AATAATTTCA AGTCATTTCT GCCTGCTAAT AGACCCACAT	9600
AGTAAATCTG AGGAATACCT GGGGCAAAAG CTGGAATTAG ACGAGCGAGA AAATACTTGA	9660
CATCATCATC TCCAAGCGCT GAATAGTAGG TTGAATTGAT TTGGTAGATA TCTAAGTTGT	9720
TATACTCGGC ACTAGAGTAC TTACGTTTGA CATTGGCTCC AACCTTATAG AGTTCAATTTG	9780
AAGCATAGTC AATCTCCTCA TCGGTCAGGA TATCCTTGAC ATCTACTACT CCAATCCCAT	9840
CATGGGTATC TAGCGTCGTA AATTGCTTCA TCGGGCTCAT CTTTAACCAC TTAGCCAAAC	9900

757

GCTCTGTTCT GGAAGTGTAA AGAGTATAAA GTGTCACCAT TGGAGAGCA AAATCATAAA	9960
CATAGTAATC ATGGTCTGCT ATTTTAAACT GAATCGAATA GTGTTTCATGA ATCTCAGGTA	10020
AAAGCTCTGT CCCATACTCA GCAGCGATAT CTCGAACTTT GTCCAATAAA TCCCAAATAT	10080
CTGGTTCCAC AAAGAAATCA TTAGTATCCA ATTTCTTCAC TGCATAAGCA AAGGCATCTA	10140
GACGAATCAA ATCACACCCA TTACTTGCCA AGTGCTGAAT GGTCTTACGG ATAAATTCCA	10200
TAGTTACTTC TTTGGTCACA TCAAGATCAA TCTGCTCCTC ACCAAAGGTA TTCCACAAAT	10260
GTTCCTACTGA ACCATCTTCA AACACAATCT CTGCTTTGG TGCACGATCC TTACGCTTGT	10320
AAATTAAATC TACATCAGAC TGTGTCGGAC GGTTCCTGG CCAAACTTA TCCCAGTTTA	10380
AAAAGAGAGC TTAAATTCAT CTGGCTTCAT GTTTTCTTG ATAGTCCTTA TAATACTTGG	10440
ATTGACGAGA AATATGATTA ATCATAAAAT CAAACATAAG ATAATATTC TCACCTAAAC	10500
GCTTCACATC CTCCCAATCA CCAAAGCTG AGTCCACTTC GTCGTAGTCA ACTGGCGCAA	10560
ATCCACGATC AACTGTTGAT GGGAAAAATG GTAAAAGGTG AACTCCTCCA ATAGCATCTC	10620
CAAAATGCTC TTCCAAATTA TCATATAAGT CTTTAAGATT ATTTCCAAG CTATCAGAAT	10680
AGGTAATCAA CATGGTTTTA TTTTGAATTG GCATCATTAC TCTCCTTTT CTAATTGAAG	10740
CCAAGTCTCA TATGATCTGG CTTTATAAAT AAAATTCATT TTAAATCTCT ATTTATCATC	10800
AAACTCGTAC TAATATAGAC TGTGATAAAC AAAGTACTAC TTTCTTGTG TCTGCATAGA	10860
ATTATCAACA AGCTAAATC TTCCTCTGTG TCAAAGACTA TAGATTCCAT GAGCTCTTCT	10920
TATACTCTTC GAAATCTCT TCAAACCACG TCAGCTTCAC CTTGCCGTAG GTATGGTTAC	10980
TGACTTCGTC AGTTTCATCC ACAACCTCAA AACAGTGTTT TGAGCAACCT GCGGCTAGCT	11040
TCCTAGTTTG CTCTTTGATT TTCATTGAGT ATTACTTCAC TGCCCCGTG CTCATTCTG	11100
AAATGATATG GCGTTGGAAG AAGAGATAGA CAATGGTGAT ACTGATAATG CCGACCACGT	11160
AAGAGGCAAA GCTTGCTCCG TAGTCCTTGA AATATTGGCC TCGGTAGTTG TATTGGAACA	11220
AAGGCAGAGT CCACATTTTG GAATCCCGGT TCAAGACAAG GAGTGGCAAC ATGAAGTCAT	11280
TCCAGAACCA AAGGGCATTG ATGATCATGG TTGTCGCATG CATCGGTTTC ATCATTGGGA	11340
AGATGATGCG GAAATAGGTT GTAAATTGAT TAGCCCCATC GATCTCTGCT GCTTCATCCA	11400
GACTTTCTGG AATCGAGATT TTGATATAGC CAACATAGAG AAAGAGGGTC TGTGGAATCG	11460
CATAGGTCAA GTAGAGCAAG ATCAAACCA AGGTATTAGC CAAACCGACT TTAATCATCA	11520
TAACCGTAAT CGGAATCATG ATGACTTGGA AAGGTACGAA GATTCCGAGG ATTAAGAGGG	11580
TATACATGAT GGTAAGGCT TTTCTTTTAC TCATATTGCG AGCGATGGAG TAGGCTGCCA	11640

758

TAGGGATAAA GATCATTACT GCAAGTAAAG ACAAGACAGT GATGACGACA GAGTTCCAAT	11700
AATAGCCTCC AATCCCATCA GCTAAGAGAC GGCTAAAGTT GTCCCATGTG AAGTTGGTTG	11760
GAAAGCCAAA GAAATTATCT ACAATATCCT TAGTGGGTTT GAAGGAACTA AAGAGGGTAG	11820
CAAGGAGCGG CACTAAAATC AGAACCGATC CTAGAATCAA TAGAATGTAT TTGCCAATCA	11880
GGGCTTTTCT TTCATCTTGT TTCATCATGC TTCTCCTCTT AAATTTCAAA TTTCTTAGAT	11940
ACTCTCAATT GGATGATCGA AATCACTACA ATTAAGAAGA ACAAGATTAC GGCAATGGCA	12000
TTGGCATAAC CGAATTGGTT GTTTTTAAAG GCATAGTTAT AAACCAAGAG CCCAAGTGAG	12060
GTGTGGCAT TGTGGGACC ACCACCGGTC ATGGCAAAGA CTTGGTCAA GGCAGTCAGC	12120
CCACCTTTTA GGGCTAGGAT AAAGACCATA GAGACACTTG GTAGCAAGTA AGGCAATTCA	12180
ATGTTCCAGA AAACCTTGCTT GCTAGTCGCA CCATCAATCC TTGCTGCCTC TGTAATCTCA	12240
GTTGGAATAG ATTGCAAACC AGCTAGGAAG ATGATGATGG GCATAGCCAC CCCTTGCCAA	12300
AGAAGGACAA AGACAGCCGC AAAGATTGCT CCCCACTTAG TCCCTAAAAG ACTGGTTTGG	12360
AAAAATTCAA TATGAAGGGC ATTTCCAATC GCTGGAAGAC CGTAGTTGAA GACTTGCTTG	12420
AAGATCAAAG CCACTGTCAA ACCAGATAAA ACAGCTGGGA AGAAGAACCA AGCACGGAAG	12480
AAGGTTTGGC CTTTGATTTT AGAATTCAAG ACACGCGCAA TGAAGATCCC GAGTGCAATC	12540
TCACCAACCA CCATGGCAAT CGCAATGATT GCGGTAAAGC CAATCGCATT CATGAATTTT	12600
GGATCCATGA AGAGGAGCTT AAAGTTGTTT AAGCCAACAA ATTTGTAGTT ATAAGTCAAT	12660
CCTGTCCAGT TGGTAAAAC TAAAAAGGCT CCTTGAAACA TCGGCACATA GAAGAAAATT	12720
GCTTGTAACA AGAGGGGGAT GACCACAAAA GCCCATGCCC AATATTTTTG TAATACTTTT	12780
TTCATAGTCT CTCTACTCCT AATCCACATC CGCTTTCATC GGGTTAAAGA AGGCATTCAA	12840
ATCATTGACC ATGCCTTGT TATCACCGGT CAAGACATAG TTCATGGTCA AGGTATGGAA	12900
GTCTGCTTCA CTGGTCCAGT ATTGTTGCAA CCAGACCAAG TGACGATCCG TAAAGGCATA	12960
TTCGGTCATA CCAGCAAGCG GTGAATCTTC TCCTGCTTGT TTGACCCCTT CGATCGCTGT	13020
TGGAGATCCG TCCACATCGT AGTATTTTTG CATGACTTCT GGACGGGTCA TATATTCCAC	13080
AAAGGCATTG GCTTCTTTTG GATGTTTGGT GGTGGCTGAG ATAGACCATG CCAAGTCTCC	13140
CGCACCAACG GTTAAGCTTT GTCTTTTTTC TTTTCTGGA ATCATGAAGG TCCCAATCTT	13200
AAAGTTCGGT TTTTGTTTAT TAATCGCTGT GATCGCCCAA GACCCATTGT GTGTCATGAG	13260
GACATCCCA CGTGCGAAGG CTCCGATAAC ATCGGTATAG CCAGCACCTT CCCAGTTCTT	13320
TTGCTTAGAT CCATTGATGC GAAGGATGTC CATGACCTTG ATATCATCTT TCATAATCGG	13380
ATCCGACAAT TTAATGGCAT TTGGTTGAGA ATAACGAAGG TATTGATTTG CTTCTTTTCC	13440

TCCACCTGTT	GCTGTCGCAA	AGGCTAATTG	ATTGTAACCA	TTGAGTGTCC	AAGCATCTGC	13500
ACCTGCAATT	CCAAATGGTG	TTTGTCTT	AGCAACGATA	TCTTTGACTA	ACTGTTCAAA	13560
TTTCATCCCAG	GTTCAGGAA	CCTTCAAGCC	CAGTTCCTCG	AATTTATCTT	TGTTGTAGTA	13620
AATTCCATAA	GCATTAGCTG	TAAAAGGAAC	GTGTAAACT	TTTTCGTTTA	CAGCATATTT	13680
TTCAGCGTAG	CCATTTTCA	CGCGTTTCAG	GTAGTCTTTG	TTGCTCAAAT	CTTCAAAAAC	13740
ACCTGCTTTT	GCCCATTCTT	GCAGTTCGAT	GGACTGTGGG	TAAATATTGA	CCACATCAGG	13800
CACATCTCCT	GCGAGAACGC	GTGTCTTCAA	TACTTCACCA	GCATTTGGTA	CATTGACGAC	13860
TTTGACCTTG	ATCTTAGGGT	TTTCCTTCTC	AAAATCACGA	GTGATTTCTT	CCAAGGTTTT	13920
GGTCATTTCT	TTTTTCTGGT	TGAAATACTC	GATGGTCACT	GTGCCATCCG	CAGATTTACC	13980
ATAGTTGGAG	CAAGCGCCGA	GCCCAAACAA	AGCTAAACCT	GTAGTTGCAA	GAAGTCCGAT	14040
TTTTTTATAC	CATTCCATTA	GAAAGCCTCC	TTTATAAATT	TATACACCCT	TATTGAACTG	14100
CACCCCAAAA	GTTAGACAGA	ATAAATCTAA	CTTTTGGGGT	CAGTACATAT	CATAGTTTTC	14160
TAAAAATATA	CTGTCTACTC	AAAAAATCTC	CTTGGGATAA	GATAACAGTT	AAGCCCGCAT	14220
ACATTAGTTC	TGCACCTGAG	TAAACTTCGC	CATTTTCCTG	TAATTTATAT	AGTCCCTCTT	14280
CATCCAAATC	TTTTAATTTT	AAAGTTGTTT	CCATGGTCTC	TACAACAGAT	AAAACGCGAA	14340
CGTAGGTTAC	AATCGTTTGA	TTCCGTAAT	TAAATTGTAC	AGCTGCTTCA	TTGGATACAG	14400
TATCAGGATT	AATTAGTCTA	TACTGCTGTC	CTAACTGAAC	TACTGGTCGT	AATTCTTTAT	14460
ACAAGTTCAC	CTGATTAGCA	ATCGTAGCTT	TCTCTTCATC	TGATAAATTT	GTCAAATCAA	14520
GTTCATAGCC	CAAATTTCCC	ATCATTGCTA	CAAGGCCACG	TGTTTCTAAT	GGTGTCAATC	14580
GTCCCATCTG	ATGATTCCGT	ACTGCTGACA	CATGAGCCCC	CATAGAAATG	GTTGGATAGA	14640
GATAGGATGA	ACCGTATTGA	ATTGGTAAAC	GTGCAATGGC	ATCAGTATTA	TCACTAGCCC	14700
AGACTTGTGG	GAAATAGCGC	ATCATACCAA	GATCATTTTCG	TCCACCACCA	CCAGAGCAGG	14760
ACTCAAAGAG	AATATGGCTG	TGCTTCTCTG	TCAGATAAGA	AACGAGTTCA	TAAAGCCCCA	14820
GCATGTACTG	ATGAGATTGC	ATCTGTGTCT	CTAGATAAGT	TAATCCATTC	CCTAGCTTAG	14880
TGATATTGCG	GTTTCATATCC	CATTTAATGT	AATCAATATC	ATGATAAAAT	AGGAGTTGAT	14940
CTAAGACACT	TTTCAAGTAT	TCTACTACCT	GAGGATTGGC	AAGATTAAGT	ACTAATTGAT	15000
TCCGAGAATA	AGTATGCTCA	TAGCCAGGAA	CCTGAATAGC	CCAGTCAGGA	TGTTGACGAT	15060
ACAAATCACT	ATCTACAGAA	ATCATTTTCGG	GTTCTAACCA	AAGTCCAAAC	TGCAAACCTC	15120
TTTCATGGAT	AGCTGAAATC	AGACTTTCTA	GACTTCCACC	CAGTTTTTCC	TCATTAACAA	15180

760

CCCAATCACC TAAAGCACGA TTATCATCAA AACGATTGCC AAACCAACCA TCATCTAATA	15240
CAAAAAGTTC AATGCCAACT TTCTTAGCTT CATCTGCTAA CTCTAACAGT TTTTCTCTCT	15300
GAAAGTCAAA GTAAGTAGCT TCCCAGTTAT TGATTAGAAT TGGACGTTCT TTTT TAGAAA	15360
ATTCACCTAG CATAATGTGC TTCAGTACAA AATTCTGACT TTCATGACTA ATACCAGTTA	15420
ATCCCTGATC TGAATGAGTC ACTAAAGCTA CCGGTGTTTC AAAGTATTCC TCAGGAGCTA	15480
ACTTCCAAGA AAAGTTTCTT GGATTAATGC CAATAGCCAC CCGAACTTCA TTCAATTGAT	15540
TTTTTTGAAC AAAAGCTTCA AAGTTGCCAC TATACATTAG TTGAATAGCA AACACATTCC	15600
CAGCATCCTC TGTGACTCCT TGTTCGCATA GTAGAAGAGC TGGTGTGTTGA GCATGACCAG	15660
AAGCACCTCG GTTTGAACTA ATCGAAAAGA TTCCTTGTTT TACCTGTTGA CGTCTAACAG	15720
TCTTTTCACG AGCATAAGCA CCCTGCAGAG TTACTATTTT GTAATCTGCA GCTGGAAAAT	15780
CAGCCATAAA AGAAAAATCT TTATGGATGA CAACTTCCTG ATTACTATTA TTATCTAATT	15840
TACTGTAGCT AGCAATAGTC GCATCATTAT TAAAAGTAGT ATAATACAAA GTCAGACTAA	15900
GTTGAGCCTT AGAATCTTCT AACATTAAGA CAAGAGTCTC TGTATCGTCC ATGCTATGTG	15960
GAGAAGGTAA GCCCTGTGGA CCATTCTGAC CTTTTAAAAT CTTTGCTTCT ACAAATCGAA	16020
AGTCTGTTAC TTCAGTTACA CTATGCTGAA CCTGTATGGT TGGTTTCCTA AAATCTCCTA	16080
AGCCATGTTG TCCAAAAATC TGTCGCTGAG TATCTAAACT AAAGGTTCTA TTAGTAGCCG	16140
TTGGATTTCC TGAAAAGGCA TGGTCTCGTT CATAAACACT ATTGGAACCT TTATAGTTCT	16200
TAATAGTCTT TCCTAAATGT TTCAAAAGTA AGTAGCCATT TCGATTTTCA ATAATCAAAC	16260
TTAGATTTTT ACTCTCAACA TAAAATAGAT TATTCTCTAT CCTAACTCCC ATTTACTTCA	16320
CCTCATCACT TTATTGATTA TATTTTATCA CCTGAAATCG CTTTCCAAAA TAGAAAAATG	16380
TCTCAAGAAT ATGGTAAAAAT GTTAGGTAGG AGGTAGCACA TGTTAGTTTT TTCAGAATAC	16440
CAGACTGGAA CAATCGACCT TGCCCTAAGC TTTTATGGAT ATGAGGAATG CACACCTAAT	16500
TACTCTTTTG GTCCAGCCAT TCGTGATACA TACGTTCTAC ATTACATTAC TAAAGGACAA	16560
GGAAAATTTT ATTACAAGGG TAAAATTGTT GATTTAAAAG AAGGAGATTT CTTTCTATTA	16620
AAACCAGAGG AACTAACCTT TTATCAAGCA GATAGTAAAG AACCTTGGGC CTACTACTGG	16680
TTAGGAATCA CTGGAGGGAA AGCCCTGAT TATTTTGCTC TTTCCCAAAT TTCTGATCAA	16740
TCCTATCTCA TCCAATCTGA AACTTGTCAT ACCCAGACTA CTGCAAACT CATCTCAGAC	16800
ATTGTCCGCT TCGCTCAGAT TACAAAATCA AGTGAATTAG CTCAACTCCA TATCATGGGA	16860
CAACTTCATG AACTGATGTT TCATCTGGGA ACTATTGCTC CCAATCAGAA AAAAAAGAAT	16920
ATTTTCATCAA CCCACCAACT CTATCTTGAA TGCAAACGAT TAATTGATAG CCACTATCCT	16980



761

CAATCACTTA CAATTCAAGA TTTAGCAAAA GAACTATCCG TTCACAGAAG CTACTTATCA	17040
AGCGTATTCA AAGAATTTAA TACCTTATCA CCCAAAGAAT ACCTACTCTA CGTTCGAATG	17100
CACCGAGCTA GACAACTTCT CGAAAATACC CAAGAGTCCA TCAAGGTAAT TGCATACTCG	17160
GTAGGTTTTT CAGATCCACT CCATTTTTCG AAAGCTTATA AACAACTT TAATCAGACT	17220
CCAAGTCATA CAAGAAAAGA ATACTCTCAA TACCAACTAG TAAGAAAGGC AACATTATGA	17280
<hr/>	
AATCCTACCA AGCTGTCTAC CAAATCCTAT CTAAAGAAAC CGACTATATC AGCGGAGAAA	17340
AAATCGCAGA AAACTATCC CTAAGCCGAA CAGCAATTTG GAAAGCCATC AAGCGACTAG	17400
AACAAGAAGG CATTGAAATT GATAGTATCA AAAATAGAGG ATATAAACTG ATGAATGGTG	17460
ACCTTATTCT TCCAGAGATT CTAGAAGAAA ATCTTCCAAT TAAAGTCAGC TTTAAACCCG	17520
AAACAAAATC AACACAATA GATGCAAAAG AAGCAATTGA TTTAGGCCAT GAAGCAAATA	17580
CCCTCTATCT AGCTTCCTAT CAAACAGCAG GCCGAGGCCG TTTTCAACGT TCCTTCTACT	17640
CACCACAAGG TGGTATTTAT ATGACACTCC ATCTTAAACC AAATCTCCCC TATGACAAAT	17700
TACCATCCTA CAACTACTT GTAGCTGGAG CTGTCTACAA AGCCATTAAG AACCTAACTT	17760
TAATAGATGT CGACATAAAA TGGGTCAATG ATATCTATCT AAACAATCAT AAAATTGGAG	17820
GAATCCTTAC TGAAGCAATG ACCTCTGTAG AAAGTGGCTT AGTCACAGAT ATCATTATTG	17880
GAGTAGGTAT CAATTTCACT ATTAAAGACT TCCCTCAGGA ATTAAGAGAA AAAGCTGCCA	17940
GCTTATTTAA AGCTACAGCT CCTATAACAA GGAATGAATT GATCATAGAA ATCTGGCGTG	18000
CTTCTTCGA AACACCAGCA GAAGAGCTAT TATACCTATA CAAAAACAG TCATTCAATC	18060
TAGGAAAAGA AGTCACTTTC AACTAGAGC AAAAAGACTA CAAGGGACTT GCTAAAGACA	18120
TCTCAGAAAA TGGAAAACCT TTAGTTCAAT GTGATAACGG AAAAGAAATC TGGCTAAATA	18180
GTGGCGAAAT TTCTCTCAAT AGTTGGAAGT AAAATAACAC AATTATAATA TAAACGATAT	18240
AAAAATAACT TCAGATTAGT AATTCAATTA AGTTTACGG ATCTGAAGTT TTATTGGCTC	18300
TAAAAATAAA AAAGAGAGTT ACAGACTCTC ATTAAAACGG AGAATAAGGG ATTCTGAACCC	18360
TTGCGCCAGT TACCCGACCT AACGATTTAG CAAACCGTCC TCTTCAGCCT CTTGAGTAAT	18420
TCTCCAATTA ATGGGCACGA GTGGACTCGA ACCACCGACC TCACGCTTAT CAGGCGTGCG	18480
CTCTAACCAC CTGAGCTACG CGCCCAAGTT AAAAAACTTG GTAATTTGAA CAAAGTTCAA	18540
AGCGGGTGAC GAGAATCGAA CTCGCGACAA CAGCTTGGAA GGCTGTAGTT TTACCACTAA	18600
ACTACACCCG CATAAATACT ATCAATAAAA TGGCGCGAGA CGGAATCGAA CCGCCGACAC	18660
ATGGAGCTTC AATCCATTGC TCTACCAACT GAGCTACCGA GCCTTATTGC GGGAGCAGGA	18720

762

TTTGAACCTA CGACCTTCGG GTTATGAGCC CGACGAGCTA CCGAGCTGCT CCATCCCGCG	18780
TTAATAATAT AAAAGGAGGA TGTGGGATTC GAACCCACGC ACGCTTTTAC ACGCCTGACG	18840
GTTTTCAAGA CCGTTCCTT CAGCCGGACT TGGGTAATCC TCCAATATTC AAATGGACCT	18900
TGTAGGACTT GAACCTACGA CCACTCGGTT ATGAGCCGAG AGCTCTAACC AGCTGAGCTA	18960
AAGGTCCGAC AAGATCATT TAGCGGCGAA GGGGATCGAA CCCCCGACCT CCCGGGTATG	19020
AACCGGACGC TCTAGCCAGC TGAGCTACAC CGCCATGAAT CGGGAAGACA GGATTGGAAC	19080
CTGCGACACC TTGGTCCCAA ACCAAGTACT CTACCAAGCT GAGCTACTTC CCGAGTTAAA	19140
TAGAAAAATG CACCCTAGAG GAGTCGAACC TCTAACCGCC TGATTCTAG TCAGGTACTC	19200
TATCCAGTTG AGCTAAGGGT GCTCCATATT ATGCCGAGGA CCGGAATCGA ACCGGTACGA	19260
TCGTTACCAA TCGCAGGATT TTAAGTCTG TCGCTCTGCC AGTTCCGCCA CCCCCGGCTC	19320
TCTAAGCGAA CGACGGGATT CGAACCCGCG ACCCCCACCT TGGCAAGGTG GTGTTCTACC	19380
ACTGAACTAC GTTCGCACTG TTTTCTTCTA TCTAAAAATG CCGGCTACAT GACTTGAACA	19440
CGCGACCCTC TGATTACAAA TCAGATGCTC TACCAACTGA GCTAAGCCGG CTCATTTGTT	19500
ATATCTTAAT GCGGGTTAAG GGACTTGAAC CCCCACGCCG TTAAGCGCCA GATCCTAAAT	19560
CTGGTGCGTC TGCCAATTCC GCCAAACCCG CATATATGAC CCGTACTGGG CTCGAACCAG	19620
TGACCCATTG ATTAAAAGTC AATTGCTCTA CCAACTGAGC TAACGAGTCT AAAATAACTT	19680
GCGTTACCTT AAACGGTCCG ACGGAATCGA CCCGGTAC	19718

## (2) INFORMATION FOR SEQ ID NO: 100:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4117 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

CCGTGGAAAA GTCTGGATAG TGAATGGTCT TCACACAATG ACCTGAAAGA AGCCTGAGAA	60
TAATTATGGA GAGTAGCATT CTGAGAGGTG TTACGAGAAC CATATGACAG AGCTGTTTGA	120
AGAGGGAATA TTGAGGAGAA AAATCCTGAG CCTACCAGTT GGAGTTGGAA AGAGCTGACT	180
GTTAGATCAT GGTATTATTAT CCACAACCTG TGGATAACTT TGTGAATAAG AGAAGTTGCT	240
AAAGAAGGAG ATATATAACG ATGAAGAAAA TCAAACCGCA TGGACCGTTA CCAAGTCAGA	300
CTCAGCTAGC TTATCTGGGA GATGAACTAG CAGCTTTTAT CCACTTCGGT CCTAATACCT	360
TTTATGACCA AGAATGGGGG ACTGGACAGG AGGATCCTGA GCGCTTTAAC CCGAGTCAGT	420

763

TGGATGCGCG TGAGTGGGTT CGTGTGCTCA AGGAAACGGG CTTCAAAAAG TTGATTTTGG	480
TGGTCAAGCA CCACGATGGC TTTGTCTTTT ATCCGACAGC TCACACAGAT TATTCGGTTA	540
AGGTCAGTCC TTGGAGGAGA GGAAAGGGCG ACTTGCTCCT TGAAGTATCC CAAGCTGCCA	600
CAGAGTTTGA TATGGATATG GGGGTCTACC TGTCAACGTG GGATGCCCAT AGTCCCCTCT	660
ATCATGTGGA CCGAGAAGCG GACTACAATG CCTATTATCT GGCTCAGTTG AAGGAAATCT	720
TATCAAATCC TAACTATGGG AATGCTGGTA AGTTCGCTGA GGTTTGGATG GATGGTGCCA	780
GAGGAGAGGG CGCGCAAAAG GTTAATTATG AATTTGAAAA ATGGTTTGAA ACCATTCTGT	840
ACCTGCAGGG CGATTGCTTG ATTTTTCATA CAGAAGGCAC CAGTATCCGC TGGATTGGCA	900
ATGAACGAGG GTATGCAGGT GATCCACTGT GGCAAAAGGT GAATCCTGAT AAAGTAGGAA	960
CAGAAGCAGA GCTGAACTAT CTTCAGCACG GGGATCCCTC GGGCACGATT TTTTCAATCG	1020
GAGAGGCAGA TGTTCATC CGTCCAGGCT GGTTCCTACCA TGAGGATCAG GATCCTAAGT	1080
CTCTCGAGGA GTTGGTCGAA ATCTACTTTC ACTCAGTAGG GCGAGGAACT CCACTCTTGC	1140
TTAATATTCC GCCGAATCAA GCTGGGCTCT TTGATGCAAA GGATATTGAA CGACTTTATG	1200
AATTTGCGAC CTATCGCAAT GAGCTCTATA AAGAAGATTT GGCTCTGGGA GCTGAGGTAT	1260
CTGGTCCAGC TCTTCCGCA GACTTTGCTT GTCGCCATTT GACAGACGGC CTTGAGACCA	1320
GCTCTTGGGC AAGCGATGCA GACTTGCCCA TCCAGTTAGA ACTCGACTTA GGTTCCTCTA	1380
AAACTTTTGA TGTAATTGAG TTAAGAGAAG ATTTGAAGCT AGGGCAACGA ATCGCTGCTT	1440
TTCATGTGCA AGTAGAGGTG GATGGTGTCT GGCAGGAGTT TGGTTCGGGT CATACTGTTG	1500
GTTACAAACG TCTCTTACGA GGAGCAGTTG TTGAGGCACA GAAGATACGT GTAGTCATTA	1560
CAGAATCACA GGCTTTGCCT TTGTTGACCA AGATTTCCCT TTATAAACT CCTGGATTAT	1620
CAAAAAAAGA AGTTGTTTCA GAACTAGCAT TTGCAGAAAA AAGCCTAGCT GTGGCAAAGG	1680
GAGAAAATGC CTATTTTACA GTTAAGCGCA GAGAATGTAG TGGTCCTTTA GAAGCTAAGA	1740
TTTCGATTCA ACCGGGGACA GGTGTCCATG GTGTGCGCTA TCAGGATGAG ATTCAAGTCC	1800
TTGCGTTTCA AACTGGTGAG ACTGAAAAAA GTCTGACGCT ACCAACCTTG TATTTTCGAG	1860
GAGATAAAAC CTTGGATTTC TATCTGAACC TAACGGTGA TGGTCAGCTT GTGGATCAAC	1920
TTCAAGTCCA AGTTTCATAA AAGAAGAACC TTTGCGCGAT GCAAAGGTTT TTTTGGTTAT	1980
TAGTGACTTG GTAACAGCT GAGGGTGAAA GTTAGTTGTT CAGCTTTTAA GAGGTCTTGG	2040
TGTTGAATAG TTGATACGAG TGTMTGTCC AGTCGGCATT CTTTGACAAA GTTAAAATGG	2100
TTGTGGTTTT GTTTAGTATG GATATCCAGC CATTTATCTT CTTTAGCGAG GTAGACTCGT	2160

764

AGATGGTCAA AGAGAGGGAT TCCGAGGTCA TAGCTTGGTT TTCCTGGACA GGTTGGATAA	2220
AATCCGAGAG CTGACCAGAT GTACCAAGCA GAGAGACTAC CATGTCTTTC ATCTCCAGGA	2280
TAGGCTTCCC AACTTGGGTG AAAAGCTTTC TGACGGAGCG TCTTGATAAG AAGGGCAGTG	2340
TAGTCAGGGT AATCGCTGTA ACGGAAGAGA TAAGGAATGT GGAAACTAGG CTGGTTGGAA	2400
ATGGCTATTT GTCCAAAAGG AGCAGTAGCC ATCTCGCTCA TTTCGTGAAT TTCGTAACCA	2460
TAGCCTGTTG TTTCAAAGAG GGGAGCATCT TGACAGGCTT TCAAAAGATA GTTGCTAAAG	2520
GTTTCTTTTC CACCCATCAG TTGGATTAAG CCAGGGATGT CGTGAGAAC GCCTAAAGTA	2580
GCTTGAATGG CAGAGCATTC AGCGTAGTCT CGCCCCAAC TATAAGGAGA GAAGTCAGGG	2640
TGAAAGTTTC CTTGATTGTC TCGTGCTCGC ATGTAACCTG TCTCAGCGTC AAATAGCTGG	2700
CGGTAATTTT GTGAAGCAGC CTTGTAGGTT TCAGCGATTT CTATGTTCTC TAGTTTTTTG	2760
GCACAGCTGG CGATACAAAA GTCACTATAG GCATAGTCTA GAGTATGGCT AACACTTTCG	2820
TGGTGGTCGG TAGAGAGGTA ACCTAGTTCT TGGTATTGGG CTAGTCCGTG GCGGCCATTG	2880
ATGCCGAGAG GGTCCGGCTTT GCTGGCTGTT TCGAGCATGG CTTGGAAGAG TTCTCCTTCT	2940
AGGTCGGGGG TCATGTCCTT GCAGGCGCTA TCTGCGATAA TACCGTCTAA AAGTGTACCT	3000
GGCATCATAC CCCGTTTCATC TGGAGCCAGC CATTTTGGAA GGAAACCACT ATCGCGGTAG	3060
CTATTGAGGA AACCTTCTAA AAAGCGTTGA TAGTGCTCCG GTATGATAAG GGCAAAGAGG	3120
GGGAAGGTGG TGCGGAAGGT ATCCCAGAAA CCATTGTTGC TAAAGAGGAC ACCAGGCTTG	3180
ACAGTACCAG TAGCCAGATC CATGTGGATG GCTTGCCCTG ATTCATTAAAT CTCATAAAAA	3240
GTCTGTGGGA AGAGGAAGAG TCTGTAGAGG CAGTGGTCAA AGAAGGTTCG GTCAGCCTCT	3300
CCTGTCTCTA TAATGTCAAA ACGATGGAGG AGATTTTCCC AATCCACTTG GGCACCTCAT	3360
TTACAGCTAT CAAAATCTTC TTGAGGTAGA TTGATTAGAG CTTGAGAAGG AGAGATGAAA	3420
GAAGTGGCTA GTTGCACTGC GGTTTGACTA CTTGCTAAGT CAATTGCGCA GTCTCCAGCT	3480
TCTTGGCTGA TAGCAAGAAT ATCCGTGTTT ATTTGCAGGG CAGTGAACAT CGTTAGCGAA	3540
TTTTTGTTAG TTTCAGTTTT ACCTTCTTGT CGCAGGGCAA GAGTCCGCTT ATCTACTTGC	3600
TCTACTGTCA GTTCATCTGC TGCCTGAAGA TAGAGGGAGA GGGCTTTGCC TTGCTTTTGA	3660
TTCAAACGAA TAGAAGCACC ATAGCAAGTC GGTGTGAGCT GGGTTTCAAT CTGATAACGC	3720
AGAGAAAAGA GCTTCAAATA GTGAGGCTGG AAGCAAGCTT TATCTATATC ATAAGAAGAC	3780
TGGCGGTGAA AGAGGCTGTC TCCCCCAGT TGACTGGTGA CAGGTGTCAG AAGGAGCCAA	3840
GAGTAGTCCC CAATCCAAGG ACTGGGCTGG TGAGTTAATC GAATCCCCTG AAAGATAGGC	3900
AGATGTGGAT CAAAAACCA AGATCCATCC TGGTCACTGG TCTGGGGCAC AAAGTAATTC	3960

765

ATCCCAAAAG GCACGCCTGT GTATGGCAGG GTATTTCCCC GAGAAAAGGC ATGCTTGTTC 4020  
 GTAGTTCCAA AACGGGTATC GATGGTATCA AGTAGTGGTT TCATAGTCTT TCCTTTAGCT 4080  
 GTTTTCTAC ATTATATCAG TAATAGAGGG CCTTTAG 4117

(2) INFORMATION FOR SEQ ID NO: 101:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2727-base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

CTGGTTCAAT TATTATTCAC TCTAAGTAGT CATATGTTCT TTATTTATGT GAGTTTTTAC 60  
 CTTTAAAGG ATCTTGTAG ATGGGAGAAG GTTTTAAAG TGACAGATGA TAATACAAGA 120  
 AAAGTTCGTT TATTAGTAGC CTTTTTAGC ATTGTCATAG GCTACATCCT GAGTTCCTTC 180  
 TTTATTAGCC TGTATCATTT GTGGCAACAA GCGCTTAGAG GATTATTATG AAATCAAGAG 240  
 TAAAGGAAAC GAGTATGGAT AAAATTGTGG TTCAAGGTGG CGATAATCGT CTGGTAGGAA 300  
 GCGTGACGAT CGAGGGAGCA AAAAATGCAG TCTTACCCTT GTTGGCAGCG ACTATTCTAG 360  
 CAAGTGAAGG AAAGACCGTC TTGCAGAATG TTCCGATTTT GTCGGATGTC TTTATTATGA 420  
 ATCAGGTAGT TGGTGGTTTG AATGCCAAGG TTGACTTTGA TGAGGAAGCT CATCTGTCA 480  
 AGGTGGATGC TACTGGCGAC ATCACTGAGG AAGCCCCCTTA CAAGTATGTC AGCAAGATGC 540  
 GCGCTCCAT CGTTGTATTA GGGCCAATCC TTGCCCCGTG GGGTCATGCC AAGGTATCCA 600  
 TGCCAGGTGG TTGTACGATT GGTAGCCGTC CTATTGATCT TCATTTGAAA GGTCTGGAAG 660  
 CTATGGGGGT TAAGATTAGT CAGACAGCTG GTTACATCGA AGCCAAGGCA GAACGCTTGC 720  
 ATGGTGCTCA TATCTATATG GACTTTCCAA GTGTTGGTGC AACGCAGAAC TTGATGATGG 780  
 CAGCGACTCT GGCTGATGGG GTGACAGTGA TTGAGAATGC TGCGCGTGAG CCTGAGATTG 840  
 TTGACTTAGC CATTCTCCTT AATGAAATGG GAGCCAAAGT CAAAGGTGCT GGTACAGAGA 900  
 CTATAACCAT TACTGGTGTG GAGAACTTC ATGGTACGAC TCACAATGTA GTCCAAGACC 960  
 GTATCGAAGC AGGAACCTTT ATGGTAGCTG CTGCCATGAC TGGTGGTGAT GTCTTGATTC 1020  
 GAGACGCTGT CTGGGAGCAC AACCGTCCCT TGATTGCCAA GTTACTTGAA ATGGGTGTTG 1080  
 AAGTAATTGA AGAAGACGAA GGAATTCTGT TTCGTTCTCA ACTAGAAAAT CTAAAAGCTG 1140  
 TTCATGTGAA AACCTTGCCC CACCCAGGAT TTCCAACAGA TATGCAGGCT CAATTTACAG 1200

766

CCTTGATGAC AGTTGCAAAA GGCGAATCAA CCATGGTGGG GACAGTTTTC GAAAATCGTT	1260
TCCAACACCT AGAAGAGATG CGCCGCATGG GCTTGCATTG TGAGATTATC CGTGATACAG	1320
CTCGTATTGT TGGTGGACAG CCTTTGCAGG GAGCAGAAGT TCTTTCAACT GACCTTCGTG	1380
CCAGTCCGGC CTTGATTTTG ACAGGTTTGG TAGCACAGGG AGAAACTGTG GTCGGTAAAT	1440
TGGTTCACCT GGATAGAGGT TACTACGGTT TCCATGAGAA GTTGGCGCAG CTAGGTGCTA	1500
AGATTCAGCG GATTGAGGCA AGTGATGAAG ATGAATAAGA AATCAAGCTA CGTAGTCAAG	1560
CGTTTACTTT TAGTCATCAT AGTACTGATT TTAGGTACTC TGGCTCTAGG AATCGGTTTA	1620
ATGGTAGGTT ATGGAATCTT GGGCAAGGCT CAAGATCCAT GGGCTATCCT GTCTCCAGCA	1680
AAATGGCAGG AATTGATTCA TAAATTTACA GGAAATTAGG CTGGAGAACC AGCCTTTTTTC	1740
TAAAGATAAG GAGAAATATG AACAAAAAAA CAAGACAGAC ACTAATCGGA CTGCTAGTGT	1800
TATTGCTTTT GTCTACAGGG AGCTATTATA TCAAGCAGAT GCCGTCGGCA CCTAATAGTC	1860
CCAAAACCAA TCTTAGTCAG AAAAAACAAG CGTCTGAAGC TCCTAGTCAA GCATTGGCAG	1920
AGAGTGCTTT AACAGACGCA GTCAAGAGTC AAATAAAGGG GAGTCTGGAG TGGAAATGGCT	1980
CAGGTGCTTT TATCGTCAAT GGTAATAAAA CAAATCTAGA TGCCAAGGTT TCAAGTAAGC	2040
CCTACGCTGA CAATAAAACA AAGACAGTGG GCAAGGAAAC TGTCCAACC GTAGCTAATG	2100
CCCTCTTGTC TAAGGCCACT CGTCAGTACA AGAATCGTAA AGAAACTGGG AATGGTTCAA	2160
CTTCTTGGAC TCCTCCAGGT TGGCATCAGG TCAAGAATCT AAAGGGCTCT TATACCCATG	2220
CAGTCGATAG AGGTCATTG TTAGGCTATG CCTTAATCGG TGTTTGGAT GGTTTTGATG	2280
CCTCAACAAG CAATCCTAAA AACATTGCTG TTCAGACAGC CTGGGCAAAT CAGGCACAAG	2340
CCGAGTATTC GACTGGTCAA AACTACTATG AAAGCAAGGT GCGTAAAGCC TTGGACCAAA	2400
ACAAGCGTGT CCGTTACCGT GTAACCCCTT ACTACGCTTC AAACGAGGAT TTAGTTCCTT	2460
CAGCTTCACA GATTGAAGCC AAGTCTTCGG ATGGAGAATT GGAATTCAAT GTTCTAGTTC	2520
CCAATGTTCA AAAGGGACTT CAACTGGATT ACCGAACTGG AGAAGTAACT GTAACCTAGT	2580
AAAAGATACG CCTACACTCC TATGTCACTT ATGGATGTAG GAGTTCTTTT TACTAGTTTA	2640
AGCAGGACTA AGACAGGTAC TAAGACAAAA TAGCAACTTC TAAACTAAC TTCCAGTTTT	2700
GGGAGAGAGA TGAAGTTAC TTTGAGA	2727

(2) INFORMATION FOR SEQ ID NO: 102:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5717 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

767

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

TTTTTTGTAG ATTAAAGTGG GGTGCAATTC CTAaaaaata AAAACAATT TTTGAAAATT	60
ATGTTAGCAG GAATTGCTTC AAATTCGATT TTATCACTTA CAGGTTTACT TGTMTTATTG	120
TTACATCGT ATAAATTGCT TGGACTCTTA TTTTTTATCA TTAAGTTAGG TATGATTTTT	180
ATTAATTCAA TTCCTTTTTT TCAGTATGAT AGTGGTATTA TTTTAAGATA CTTGAATTCT	240
AACAATAATA ACTTGAATTT TCAATATATA GTTCAACTTT TAATAGCATT TGTTATTATT	300
TATTTTCCTT TGAGTCAACT ATTACAGTTT TTGACACCCA ATATTATTGT TCGTAGTATA	360
GGAGGGGTGG TTGTTTCTAT ACTGCTTCT ATATTATATA TGATAGGAAG GACGAAATAT	420
GTTCTACGTA AATAGTTATG TTTTGTCTTA TAAAAAGAA GGTATAATGT ATTTACGTGG	480
TGGAGTATG CGGAAATAG CTATAGAACC TCAATTTCC CAAGAATTTA TCAACGATCT	540
ATTTAATAGT TGTAAGGAAC TATTAGAGAT AGAAGAAGTA TTAGGCAGTA AACTAACATT	600
TGAACTATAA ATGAACAAAT TTTAATTTCC GATGAGATAG ATATTGATAG TAGATATTCT	660
AGAACTAAAG GTTACTATTC GTTATTTTAT AATGAAGAGT ATAATAAAAT ACAGAATAAA	720
ACAGTATTAG TATTAGGAGC AGGAGTCTTA GGATGTTATA TATCTCTAAG TCTAAGTATG	780
TATGGAGTGA GGAACTTAT TGTGCTGAT TACGATATAA TAGAACCATC AAATTTAAAT	840
AGGCAAAATC TTTATACAGA GTCGGATGTT GGTAAGGAGA AGATTAATGT TCTTCTGAA	900
AAAATACACA AGTATAATTC AGATGTTTAC GTAGTACCTA TTTCTATTAA AGTTTCTTCA	960
GTAGAAGAAT TAGAAAAAAT TGTGCGGAA TATGGGACTA TAGATTTTAT CTTAAAGCA	1020
ATTGATACGC CCATTGATAT TATAAAAATT GTCAATCAAT TTGCTGTATC GCATAAGATA	1080
TCCTACATAT CAGGAGGGTT TAATGGATGC TATCTTATTA TTGATAATAT ATATATCCCT	1140
ACCATCGGTT CTTGCTTTGG TTGTCGGAAT ATAAACAAAG ATATAAATA GTACACTTTA	1200
TCTGATAAGA CAAAGTGGCC GACTACACCA GAGATGCCTG CTATTTTGGG AGGGATAATG	1260
ACTAATTTAA TAATTTAAAT ATTTCTGGGA TGTTATAATG AAATCCTAAT AGATAACGCT	1320
TACGTTTATA ATATGAGAAA TCATGCTCTA AGTCAAGAAA AATATGTTCT GGAAAACGGA	1380
GAATGTCCAA TTTGTAaaaa AATAATAAAG TGAAAGATAA CAATATTAGA GCGAAAACAT	1440
TTATTCGTTT AGTTTGTITT TGCTTATTAT CAGGAGGAGT AGCTTTTTTA TCTGCTATTG	1500
GGCAGTTCAC TGTATAGAA ACACAATTAA TAGTATTGTT CTTGGGTATT ATTTTGTCTA	1560
TATATTATGC TACTACAAT AAAAATATTC AAACATCATT GGAAAATATA GTATGGCTTT	1620

768

TTTCATCGTT TGAGATTTTA TTTTGTCTTG TTAATTTTAG AACATTTATT CAGTTACCAG	1680
TGGATATTTT TATTGGTATG ATAATATTTT TAATGCTGTG GATATTTATT ATGTTAGGTA	1740
TAGTGTGTCT TAGTTATTAT ATAACTTTAT TATTTAGCAA GGAGGCTTAG TATGTTTAAA	1800
AAAATAGGTA TAATGAGCAT TTGCATATAT ATAATTATTT TATACTGCTT GAGAATGTAT	1860
CGTATTATCA ATAATATTGA AACAACTTGT CTAACGGTTA TATGCTTAAT GTTATTGTTT	1920
TTTTTAAGAC GTTTATTTGA TAAAGATAAG TAAATAGATG TTAAGTAAAA ATGTAGAATA	1980
TAAAGGAGGT GCAATGAGTA TGATTGAAGT TAGCCATTTA TCAAAAAGTT TTGGTGATAA	2040
AATAGCTTTA AATAATATAA GCTTCACTGT TAAAGAAGGT TAGATTTTTG GATTTTLAGA	2100
ACCATCTGGT TCTGGAAAGA CCACAACGAT TAATATTCTG ACTGGGCAGT TCCTTGCCGA	2160
TAAAGGACAA TCTATTATTT TGGGACAAAA ATCTCAAAAT TTAACAAGCG GTGAATTAAA	2220
GAGAATTGGA TTGTTAGCG ATACAAGTGG ATTTTATGAG AAAATGTCTC TGTATAACAA	2280
TCTTCTTTTT TATAGTAAAT TTTATAATAT TAGTAAATCA CGTGTGATA ATTTGTAAA	2340
GCGAGTAGGA TTATATGATA GTCGCAAGAT GGTAGCAGGA AAATTATCCA CTGGAATGAG	2400
GCAACGAATG CTTTTCAGCAC GAGCTCTTAT CAACAACCCC GCTGTACTCT TTCTGGATGA	2460
ACCGACCTCA GGTCTAGATC CCACAACCTC TCGAACAATT CATGAGTTAA TTTTAGAATT	2520
GAAAACAGCA GGGACAACGA TTTTCTAAC GACTCATGAT ATGAATGAAG CAACTCTTTT	2580
ATGTGATTAT GTTGCCTTAT TAAATAAAGG GAAATTAGTT GAGCAAGGAG CTCCTTCTGA	2640
ACTCATTCAA AGATATAATA AAGATAAAAA GATTAAGGTT ACAGATTATA ATGGGAATCA	2700
GATAACTTTT GATTTTACAT CACTAGAACA GGTATCTCAG ACTGATCTGG AAAATATTTT	2760
TTCAATTCAT TCATGTGAGC CTACTTTAGA AGATATTTTT ATCACATTAA CAGGAGGAAA	2820
GCTAAATGCT TAAACGGTTT CTGGCTTTGG TATGTTGCG TTGTCAAATC ATCCTTTCCA	2880
ATAAGAGTAT TTTATTGCAA GTTTTAGTGC CTTTGTCTT CACATATTTT TATAAATATC	2940
TTATGGAAAC ACAGGGGAAG GTCAACGATC AACAGGCATT AGTTCTTTTG ATGATGTGTT	3000
TACCTTTTTT TTTTCTTTG GCTGTTGGAA GTCCTATAAC TATTATCTTG TCTGAAGAAA	3060
AAGAAAAGTA CAATTTACAA ACTCTTCTGT TGAGTGGTGT TAAAGGCTCC GAATACATTT	3120
TATCAACTAT GTTCTTCTCT TTTTGTCTAA CTTTGTGAT TATGGGAACT ACTCCTCTTA	3180
TTTTAGGAGT TACAATTGTA CATACTTTTA ATTATATTAC AATCGTTCTT CTAACCTCTT	3240
TATCCATCAT TTTATTCTAT TTATTGATAG GTTTAACCGC GAAGAGCCAA GTAGTAGCTC	3300
AGGTTATCAG TCTTCCTGCT ATGATTTTAG TTGCTTCTT ACCGATGCTA TCTGGTTTGG	3360
ATAAGACAGT TGCGAAGATA ACAGATTATA GTTTTATGGG ACTATTTACT AAGTTTTTCA	3420



CAAAATGGGA GGAATTTTCA TGAATAAAAA CTCTAATTCC TAATCTAACA CTACTTATTT	3480
GGATTGTTCT TCTATTAAC TTAATTACGA TAACTATTAG GAAAAAGAAA ATTTCTTAAT	3540
TGAGTTATTT TAATGATTAT AACACAAGT GGGAAGGAAA AAATGAAC TGATTTTGA	3600
CAGCAATTCT ACAGAATAGT CTTATTGCTA TATTTTGATT TGAGTGACG AAAAAAGAAA	3660
AATAACAATA GTGCTCATA TAATTGCAGA AGTTTTGGGT GATAAGATAA CTGATAAATT	3720
GCAATAAAAA ATGCAACATT TTAAATCTC CTCTATAAGT GCTTCAAAAA GTGCTTCAAA	3780
ACCTGTCTTG TAATCCAAGT ATTTTGGGG ACGGTGATTA ATAAGCTAGC AAAGCATCAT	3840
TAAGGATTTT TTCGGTAATT GTTGCCAAAT CGGTTAAGA AAATACTCAC GAAGAAGTCC	3900
ATTGCGATTC TCATTACTTC CCCTTTGCCA AGATGAATAG GCATCCGCAA AATAAACAG	3960
AATTCCCAT TGTTCATTA AAGGGTAACA AGCAAACTCT TTTCTCTGT CCGAAGTGAA	4020
AGTCTTTAAC TATTCTTTG GAAAGAGTCT TGTGAGGTCT TCAATAGCAG TCAACATGGA	4080
TTAGCTGTT TTTACTTGAC AAGTGCTAGT AGAAATAATA GAATAGTAAA AAACCTTTAA	4140
AGCAGTCCAG AGAGGCAGCT AAGGTTAGAC GGTGAAAGGG TGGAGACTAC CCATTTTTCG	4200
TGGAACCTTG CTGTTGGCAG GTTCCTTTT TCGTGGCTTC TGTGCGCCAG ACTCTCTCAC	4260
TAGTAAAGGT AAAAGGAGAA ACCTATGCCA GAACATCGTC CAATCATTGC TCTTGATTTT	4320
CCTAGTTTTG AGGCGGTCAA GGAATTTTGA GCTCTTTTCC CAGCAGAAGA AAGCCTTTAT	4380
CTCAAGGTAG GGATGGAGCT TTATTACGCA GCGGGGCTG AGATTGTGTC CTACTTAAAA	4440
GGTTTGGGTC ATAGTGCTT TTTGGATCTC AAATTCATG ACATTCCTAA TACAGTCAAG	4500
TCAGCCATGA AGATCTTGTC TCAGCTTGGT GTCGATATGA CTAATGTCCA TGCGGCTGGT	4560
GGTGTAGAGA TGATGAAGGC GGCGCGTGAA GGTCTTGGGA GTCAAGCCAA ATTGATCGCT	4620
GTAATCAGC TCACATCAAC GTCAGAAGCT CAGATGCAGG AGTTTCAAAA TATCCAAACC	4680
AGTCTGCAAG AGTCTGTGAT TCACTATGCC AAGAAGACAG CTGAAGCTGG CTTGGATGCT	4740
GTGTTTGCT CGGCTCAGGA AGTACAAGTC ATCAAGCAGG CTACCAATCC AGATTTTATC	4800
TGTCTGACAC CAGGGATTCG TCCAGCTGGT GTTGCAATTG GAGATCAAAA ACGAGTCATG	4860
ACACCTGCTG ATGCCTATCA AATCGGCAGT GACTATATCG TAGTGGGACG TCCCATTACC	4920
CAAGCTGAGG ATCCTGTTGC AGCTTATCAT GCCATCAAGG ATGAATGGAC ACAGGACTGG	4980
AATTAAAGAA CTAGATTAGA AAAATAAAAG GAGAATACCA TGACACTTGC TAAAGATATC	5040
GCTAGCCACC TCTTGAAAAT CCAAGCCGTT TACCTCAAAC CAGAGGAACC CTTCACTTGG	5100
GCATCTGGTA TCAAGTCACC GATTACACT GATAATCGTG TGACACTAGC CTATCCAGAA	5160

770

ACTCGTACCC TAAATTGAAAA TGGTTTTGTG GAAGCTATCA AAGAAGCCTT TCCTGAAGTA	5220
GAAGTGATTG CAGGAACTGC AACAGCAGGG ATTCCACACG GAGCCATTAT TGCTGATAAG	5280
ATGGAAGTGC CTTTTGCCTA CATCCGTAGT AAACCAAAAG ACCACGGAGC TGTAATCAA	5340
ATCGAAGGTC GCGTAGCTCA AGGTCAAAAA ATGGTAGTGG TTGAAGACCT TATTTCAACG	5400
GGTGGTTCAG TTCTTGAAGC TGTAGCAGCA GCCAAGCGAG AAGGAGCAGA TGTACTTGGA	5460
GTTGTAGCGA TTTTCAGCTA CCAATTGCCA AAAGCAGATA AGAACTTTGC AGATGCTGGT	5520
GTTAAACTTG TGACGCTTTC AACTATAGC GAGCTTATCC ATCTAGCCCA AGAAGAAGGT	5580
TACATCACGC CAGAGGGCCT TGATCTTCTA AAACGCTTGA AAGAAGACCA AGAAAATTGG	5640
CAAGAAGGTT AGGTCAGTAA GATAAAGAGA GACGAGGCTA CCGAGTCTCT TTTACCATT	5700
TATTTAAAAAT ATGACAG	5717

(2) INFORMATION FOR SEQ ID NO: 103:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5558 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

CCTGGACTTT CTAAATGAA ATCTTGGCAG CTGGATCAAG CCCTTCATGA GCATTTTTC	60
GAAGAAGAA TAGCTGGTCA CTTTCATGTC CTTCTATGGA CTTTTTTTAC AATGGCATTG	120
CTATCACACC CAATACCTAT CTAAGCGCCT GGTTCCGTAAA CTTTATTGCA GCTCTTCCTC	180
TAAATTTCTT AATTGTTGAA CCAATTGCCC GTTTTATACT AAGTTCTTTT CAGAAACCAT	240
TTACTGGGGA AGAAGTTGAA GATTTTCAAG ATGATGATGA AATCCCAACT ATTATCTAAG	300
CCAGTTCTGT AACTACTAA TATTTGAAAT CCACCTCCTT TTAGGGTGCA ATGGTTATAA	360
ATGAATTTT GAGAGGATCA GAATGAAAA ACTAGCAACC CTTCTTTTAC TGTCTACTGT	420
AGCCCTAGCT GGGGTAGCA GCGTCCAACG CAGTCTGCGT GGTGATGATT ATGTTGATTC	480
CAGTCTTGCT GCTGAAGAAA GTTCCAAAGT AGCTGCCCAA TCTGCCAAGG AGTTAAACGA	540
TGCTTTAACA AACGAAAACG CCAATTTCCC ACAACTATCT AAGGAAGTTG CTGAAGATGA	600
AGCCGAAGTG ATTTTCCACA CAAGCCAAGG TGATATTGCG ATTAACTCT TCCCTAACT	660
CGCTCCTCTA GCGGTGAAA ATTTCTCAC TCACGCCAAA GAAGGCTACT ATAACGGTAT	720
TACCTTCCAC CGTGTCATCG ATGGCTTTAT GGTCCAAACT GGAGATCCAA AAGGGGACGG	780
TACAGGTGGT CAGTCCATCT GGCATGACAA GGATAAGACT AAAGACAAAG GAAGTGGTTT	840

771

CAAGAACGAG ATTACTCCTT ATTTGTATAA CATCCGTGGT GCTCTTGCTA TGGCTAATAC	900
TGGTCAACCA AACACCAATG GCAGCCAGTT CTTTCATCAAC CAAAACTCTA CAGATACCTC	960
TTCTAAACTC CCTACAAGCA AGTATCCACA GAAAATTATT GAAGCCTACA AAGAAGGTGG	1020
AAACCCTAGT CTAGATGGCA AACACCCAGT CTTTGGTCAA GTGATTGACG GTATGGATGT	1080
TGTGGATAAG ATTGCTAAGG CCGAAAAAGA TGAAAAAGAC AAGCCAACTA CTGCTATCAC	1140
AATCGACAGC ATCGAAGTGG TGAAAGACTA CGATTTTAAA TCTTAAAAAC CAAAAAATA	1200
CAGTATCCAC ATTGGTACT GTATTCTTT TACTCTCAT CTTAAGTTAA ATTATTAAAA	1260
TCCCATATTT GGTCTATCCA GCCTTCATAA AAGTCTGGCT CGTGGCAGAC CATAAGGATA	1320
GATCCCCTAT ATTCTTTGAG AGCGCGTTT AGCTCATCCT TTGCATCCAC ATCCAAATGG	1380
TTGGTCGGCT COTCCAGCAC TAAAACGTTG TTTTCACGAT TCATCAAGAG ACAGAAACGA	1440
ACCTTGCTT GCTCTCCCC TGATAATACT TGAATCTGGC TTTCAATATG TTTGGTTGTC	1500
AAACCACAAC GGGCAAGGGC TGCACGGACT TCTGCTTGAT TAAGGGCAGG AAAGGCATTC	1560
CAGACAGCTT CAAGAGGAGT TTGGCGATTA CCGCCTTCTA CTTCCTGCTC AAAATAACCA	1620
AGTTCTAAAT AATCTCCAG CTCCACTTCC CCAGCGATTG GCGAGATAAT GCCCAAGAGA	1680
CTCTTCAAGA GAGTTGTTTT TCCAATACCA TTAGCACCAA TAATCGCAAC CTTTTGATTG	1740
CGTTCGAAGG TAAGATTTAA AGGCTTAGTA AGAGGACGGT CGTAACCAAT TTGCAACTTC	1800
TTGGCTTGA AGATAAAGCG CCCTGGTGTA CGAGCTGGTT TGAAATCAAA GGATGGTTTT	1860
GGTTTCTCAC TTTGGAGTTC GATAATATCC ATCTTATCCA ATTTCTTTTG ACGAGACATA	1920
GCCATATTAC GAGTTGCAAC ACGGGCTTTA TTACGAGCCA CAAAGTCCTT GAGGCTGCA	1980
ATCTCTTTCT GCTGGCGTTC GTAGGCTGCC TCTAGCTGAG ATTTCTTCAT AGCATAAACT	2040
TCTTGGAAGT GGTAGTAGTC ACCAGAGTAA CGCGTCAGCT GTTGATTTTC CACATGATAG	2100
ACAATATTAA TAACGTCATT GAGGAATGGA ATATCGTGCG AAATGAGAAC AAAGGCATTC	2160
TCATAGTTTT GGAGATAGCG CTTGAGCCAA TCAATATGCT CAGCATCCAA GTAGTTGGTC	2220
GGCTCGTCCA ACAGCAAGAT ATCAGGCTTT TCAAGGAGAA GTTTTGCCAA AAGCACCTTG	2280
GTTCTTTGCC CACCTGACAA AGAAGTTACA TCCGTATCCA TGCCAAAGTC CATAACACCA	2340
AGAGCACGCG CTACTTCGTC AATCTTAGCA TCCAAGGTAT AGAAATCACG ACTCTCCAGA	2400
CGGTCTTGAA GTTCTCCTAC TTCTTCCATG AGAGCATCAA CATCCGCGCC GTCTTCAGCC	2460
ATTTTCATAT AGAGGTCATT GATACGAGCT TCAGCTTTGA AAAGCTCATC AAAAGCCGTA	2520
CGGAGAACAT CACGCACCGA CTGTCTTTCA GCAAGGACAG AGTGCTGATC CAAGTAACCA	2580

772

GCCGTCACAT ATTTGGACCA CTCAACCTTT CCTTCATCTG GCAGCATTTT ACCAGTCACG	2640
ATACTCATAA AGGTTGATTT TCCTTCACCA TTGGCACCGA CCAGGCCGAT ATGTTCTCCC	2700
TTGAGGAGAC GGAAGGACAC ATCTTCAAAA ATTGCACGGT CACCAAAACC GTGACTCAGA	2760
TTTTTAACTT CTAAAATACT CATTTTAATT CCTTACCTTG TTTTATGTA ATCGTTTATA	2820
AAGGAGCCAA GCCAGATAGC CACCCAAAGT GTTGGTCCAC AAATCATCAA TCTCAAAGAC	2880
GCGATTGAAA TCAAAGAAAA AGTCCAAGAT TAATTGCGTA CACTCGATTG CAAGACTCAC	2940
AAGAAAACTA AAAAGAAGGA CCTTTTTTGT TTTCCGCAA TTTGGAAATA GATAAAGGAG	3000
TTGAAAAATC AGAGGAAAAA ACAAGAAGAC ATTGAGGATA TTTTGTAATA AAATCCAACA	3060
TAATTGTCCA ATGTCACTCA CTTCGCCCAG TTTCCAGAGA GAATTGAAAG GAGTCAAAAG	3120
AAAAACCAGG CGTCCAAGAT GCTGAATACC TGGAGTTCCC ACTCCACGG TAGATTGTTT	3180
TTGAGGAGTA AAGCAAAAAC AGACAATGCA AATGCTATAG AAAATGACTC CCCAGACCAA	3240
AATATGATTA TAAGTCTTCT TCATCATTAA GGATTTACCG CTGCGACTGC CTTCTGGCGG	3300
TCACGTTTCA TTGTGTTAGA GCGCAATTGT CCACAAGCTG CGTCAATATC TGTACCATGC	3360
TCTTGACGAA CCACACAGTT GACCCCTTTT TTCTTAAGCG TATCATAGAA AGCCAACACG	3420
CACTCTTTGG GACTACGGCT ATATTGGTCA TGCTCACTAA CTGGGTTATA AGGAATCAAG	3480
TTTACATAAG ACAATTTCTT GATGTTCTTG AGCAATTCAG TCAATCCAA GGCTTGTCT	3540
ACACCGTCGT TGAATTCATT AAGCATGATA TATTCAAAGG TTACACGACG GTTTGTTGTC	3600
TCAATGTAGT ATTCAATAGC AGCAAAGAGT TTTTCAATCG GAAAGGCACG GTTAATCTTC	3660
ATGATACTTG AACGAAGTTC ATTGTTAGGT GCGTGAAGAG ACACGGCAAG ATTGACCTGA	3720
ACCCCTTCAT CAGCAAAGTC ACGAATTTTA TGAGCCAAAC CTGAGGTTGA AACCGTGATG	3780
TGACGAGCAC CGATAGCCAT TCCTTTATCA TCATTGATAG TACGAAAGAA ATTCAAGACA	3840
TTGTTGTAAT TATCAAAGGG CTCACCGATT CCCATGACAA CGATATGGCT GATGCGTTCA	3900
TCCTGACCAC GCTCATCAAA GTATTTCTGA ACCAGCATGA TTTGCGCTAC GATTTACCG	3960
TTATTGAGGT CACGTTGCTT CTTAATCAAA CCAGAGGCAC AGAAGGTACA ACCGATATTA	4020
CAGCCGACCT GAGTGGTCAC ACAGACAGAT AAACCATAGT GTTGACGCAT GAGTACAGTC	4080
TCAATTAACA TACCGTCGGG CAATTCAAAG AGATATTTGA CTGTACCATC AGCAGACTCT	4140
TGCACAATAC GTTGTTCCTT GGGATTGACC ACAAAGTGGT CATTGAGCTT AGCAATCAAA	4200
TCCTTGAAA GGTGGTTCAT TTCTTCAAAT GACTGCACAC GTTTACGGTA GAGCCATTCC	4260
CAGATTTGAT CTGCACGGAA TTTCTTTTCT CCCTGCTCCA ATACCCATTC CTGCATGGTT	4320
TGATGTACCA AACTATGAAT TGAGGGTTTC ATTTCTTCTC CTTATTCTCT ACTCACTTCT	4380

773

GACGAATGAC AAAATGACGT TGTCCCTTGT CGTCTTTCTG ACGACGTCTA TTTTCTTAT	4440
CTGCATTCGA CTTTCGTTTA GTTTGAGTCG GTTCTTTCC TTTCTAGAA GGTGTTTCTT	4500
CTTCCGTCTT ACGCATTTTC TTGTCAAATG ATGCTCGCTT AGGGGCTTCA TTTTCTAAGA	4560
CAAAATAGGC ACAACCATAA CTACAATACT CTAAGGTA GTCTTGTAAG CGACTGATTT	4620
TTTCAAGTTT TTCTTCTGTT CGGTATCCT TGTAAAAACC TCGTAGGCGA AGCTGTTCGT	4680
TGCTCCAGTC CCCCACGATA TAATCAAAC TGGTTAATAC TTCTGAAAA CGCTGATTAA	4740
AAGTCGTAC ATCAAAGGCA TCCTTGATAT TTTCAACCA GGAAGGCT ATCCCTCCG	4800
TTTCGACCTT GTCCCGTGT AAATGGAAC CCGGACCAGG AAAGTTGTTA TAGTTGTATA	4860
ATTCAGGTGC AATTTCTTTT CGCATAGATA TCCTTTTTTC ACGATTACTT AATACTTTAT	4920
TCTACCATAA TTTCTAGCAG TTAGCAGCTT TCTCATAAAA ATGAAAAAG TCTGACGATT	4980
TTGTGAGACC AGAATCTTAT AACCTAAAA GAGAAGAACA ATTCTTCCCT CCAACTATCA	5040
TTATTTAGCA GCTGCGTACA ATTCATCTAC TTTATCCAG TTGATTACTG AAAAGAAAGC	5100
TTTGATGTAG TCAGGACGCA CGTTCCGTA TTACAGTAG TAAGCATGTT CCCAACGTC	5160
CAAGCCCAAG ATTGGTTTTT TACCTTCTGA GATTGGTGTG TCTTGGTTTG CTGTTGAAGT	5220
CACCTCAAGT TTCCCTTCTT TGTGACAAC CAACCATGCC CAACCTGAAC CAAAACGAGT	5280
TGTTGCTGCT GCAGTGAAGG CTGCTTGAA TTCTTCAAT GAACCAATG TTGCATCGAT	5340
TGCTGCTGCC AGTTCTGCTG AAGGAGCTGT TTTCTCGGA GTCATCAATT CCCAGAAAAG	5400
AGCGTGGTTC AAGTGTCCGC CACCATGTT GATAAGTCT TGACGGATAT CAGCTGGAT	5460
AGATTCTACA TCAGCAAGCA AGGCTTCAAG GTCTTACCG ATTTCAGGT GTTTTCTAA	5520
AGCTGCATTG GCATTGTTGA CATAAGTTG ATGGTGT	5558

(2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6735 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

GGAATTGTAA ATATCATATT GTTTTGCAC CCAATATCG TCGTCAATC ATTTATGGCA	60
GATACAAAGC TAGTATCGGA AGAATCATAC GTGACTTATG TGAGCGTAAG GGTGTAATAA	120
TCCATGAAGC GAATGCTTGT TCAGACCATA TTCACATGCT TATCAGTATT CCTCCGAAAC	180

774

TTAGTGTTTC GTCCTTTATG GGCTATTTAA AGGGCAAGAG CAGTTTGATG ATTTTGTATA	240
AGCATGCGAA TTTAAAATAC AAATATGGCA ATCGCAAGTT TTGGTGTAGA GGCTATTATG	300
TAGATACGGT AGGCCGTAAT CAGAAAAGTA TAGCTGAATA TATTCAGAAT CAATTACAAG	360
AAGACAGAGT AGCAGACCAG CTCACGTTAT TCGAGTCAGT AGATCCGTTT ACTGGCGAAA	420
TAAATAAGAG GAAGTAACTA AGGTGCTTTA GCACCTGCTC GGGAAAAGTG TCGCGGAGGA	480
AGCTATTTTC GTGGGCCTTT GGCCCTGGCC GGTAGAAGCG GCTTATAGCC GCAGAACAAA	540
CCACCAGTTC ACACTGGTGG TTTTGATTTA AAAAAGTTGA TACATAAAAA TAAAAGTCTA	600
TATAAAGGAT GGTAAGATTC CTGTTGTCCG ATTTGGACAA TATCCTAAAT AGTTACAATA	660
TATGGTCTAT ACTTTTCTT AGGAGAAAGC TAGATGTACA GACGTTTGAG AGATTTGAGG	720
GAGGATCATG ATCTGCCCCA AAAGCAAATA GCTACAATAC TTTCGTTTAC AAATTCAGCT	780
TATGCCAAAA TTGAACGGGG TGAGCATGCG TTGACGGCTG ATGTATTGGT TAAACTCTCA	840
GATTTCTATG ACGTCAGTAC AGACTATTTA TTGGGATTAA CTGATTTTCC TGATAAAATT	900
CGCTTTAGAA AATAATCTCC TCAATTTTCA AGAGTTTGAA AATGAGTGAG ATTTTTTATT	960
TGCCCTTTGA CAACTGAATA GCCTAAAATG GTACTTTCCT CATTTGTGGA GCAAATTTGA	1020
ATGGCTCGCC ATGATAAGAG CGATTTTAAA ATCATCAATA AAATAGAGCG ATACTTTATA	1080
TGCCATGATA CAAATGATAT ACAATGATAC TTCTGACCGT TCAGCCTGCC AACGTAAAAAG	1140
AGCAGCAAGT GAAATTCTTA TGATGACTTC ATCAGTCATG CCACGTTGAA TGTGTGAGTT	1200
TGTTAGATAA ACGCAATTAA TCCTCAAAAG GTTCCCCGAA CCTTTTGAGT TCTACAGACG	1260
CATCAGTGG AGTGTGTAAG CTTGTTGCTA AAAGCGTAAA AACCTTGGA CGAAAGGAAT	1320
AATAGACTTT CTGCGAAACA AAAATATAAT ACAATAAAC TATGAATGAT GAAGCAAGTA	1380
AACAATTGAG CGATAGCCGT TTCAAGATCC TTGTAGGTGT TCAGCGCACG ACTTTTGAAG	1440
AGATGTTAGC TGTGTTAAAA ACAGCTTATC AACGTAAACG CGCAAAAGGT GGACGAAAAA	1500
GCAAATTAAG CCTAGACGAT CTCCTTATGG TAACTATTCA ATACATGCGA GAATAGAGCA	1560
CTTATGAACA AATTGCGGCT GATTTTGGCA TTCACGAAAG CAACTTAATC CGTCGGAGTC	1620
AATGGGTGA AGCAACTCTT ATTCAAAATG GTTTTACGAT TTCAAATTTCT GCCTTAATTC	1680
TGTAAAAACA GTAAAATTCG AAGGATTGTA AGGTAAGAGT TTTTTTCTT CTGAAAAAAT	1740
GGTATAATAG CAATCAAAAC TAGAAAATAA AACGGAATTT GGAACAGATT TGTCTGTATC	1800
CTAGTAGAGT GGTGATACTA TGAAGATTAG TAAGAGGCAC TTATTAAATT ATTCCATCTT	1860
GATTCCCTAC TTGCTTTTAT CTATTTTGGG CTTGATTGTG GTCTATTCTA CCACCAGTGC	1920
TATTTTAATT GAAGAAGGCA AGAGCGCCTT GCAGTTGGTT CGAAACCAAG GAATCTTTTG	1980

775

GATTGTTAGT TTGATACTGA TTGCCTTAAT TTATAAATG AGACTAGATT TTTTGAGAAA	2040
TGAGCGACTA ATCATTTTAG TTATATTAAT AGAAATGCTT TTATTGTTCT TGGCTCGTTT	2100
TATTGGTATT TCCGTAAACG GGGCATACGG TTGGATTTTCG GTTGCAGGAA TAACTATTCA	2160
GCCAGCTGAG TACTTAAAAA TCATTATTAT TTGGTATTTA GCTCACCGAT TCTCCAAACA	2220
GCAAGAAGAA ATAGCTACTT ATGATTTTCA AGTTTGTACT CAAAATCAAT GGCTTCCCCG	2280
TGCTTTTAAT GATTGGCGAT TCGTTCTCCT AGTTCTGATT GGAAGTTTGG GAATTTTCCC	2340
TGATTTAGGA AATGCGACTA TTTTAGTCTT GGTTCCTTG ATTATGTATA CAGTTAGTGG	2400
AATCGCTTAT CGCTGGTTTT CAACCATCTT GGCCTCGTA TCTGCCGCTT CTGTCTTTGT	2460
CTTGACCACT ATCAGCCTAA TCGGTGTTGA GACCTTTTCA AAAATTCAG TATTCGGCTA	2520
TGTAGCCAAG CGCTTTAGTG CCTTTTAA TCCTTTTGCC GATCGTGCTG ATGCAGGTCA	2580
CCAGTTAGCT AATTCTTATT TTGCCATGGT CAATGGCGGT TGGTTTGGTC TAGGCTTGG	2640
AAACTCGATT GAAAAACGAG GTTATTTGCC AGAAGCTCAT ACAGACTTTG TCTTTTCTAT	2700
CGTGATTGAA GAATTTGGCT TTGTTGGTGC CAGTCTTATT TTAGCTCTCT TGTTTTTCAT	2760
GATTTTGCGG ATTATCTTGG TCGGTATCCG AGCGGAGAAT CCTTCAATG CCATGGTTGC	2820
ACTCGGTGTC GGAGGGATGA TGTTGGTTCA GGTATTTGTC AATATCGGAG GGATTTCGGG	2880
CTTGATTCCA TCTACAGGAG TGACTTTCCC CTCTTATCC CAGGGTGGAA ATAGTCTTCT	2940
AGTCTTATCA GTGGCAGTAG CCTTGTCTT AAATATTGAT GCCAGTGAAA AACGCGCTAA	3000
ATTGTACCGA GAATTGGAAA ATCAACCAAT GAACCTTCTG TTGAAGTAGG ATAAAGAAAAG	3060
GATAGTTTAT GTCTCTTCAA AAATTAGAAA ATTATAGTAA TAAAAGTGT GTGCAAGAAG	3120
AAGTCTTGAT TCTAACAGAA TTAGTGAAG ATATTACTAA AAATATGCTT GCCCCACAGA	3180
CCTTTGAAAA AATAATACAG TTGAAAGAAT TATCAACGCA GGAAGATTAT CAAGGTCTAA	3240
ACCGTCTAGT GACTAGCTTA TCAATGATG AAATGGTCTA TATTTACGC TATTTCTCTA	3300
TCTTGCCTCT TTTGATTAAT ATTTAGAGG ATGTGGATT AGCTTATGAA ATCAATCATC	3360
AAAATAATAT TGATCAGGAC TATTTAGGTA AATTATCTAC AACGATTAAA TTGGTAGCAG	3420
AAAAGGAAAA TGCCGTGAG ATCCTAGAAC ACTTGAATGT TGTCCCTGTT TTGACAGCCC	3480
ATCCAACACA AGTGCAACGC AAAAGTATGT TGGATTTAAC AAATCATATT CATAGTCTTT	3540
TGCGTAAATA CCGTGATGTT AAGTTGGGGT TGATCAATAA AGATAAATGG TACAATGATT	3600
TGCGTCGTTA CATCGAAATT ATCATGCAGA CAGACATGAT TCGTGAGAAA AAATTAAG	3660
TGACTAACGA AATCACGAAT GCTATGGAAT ATTATAACAG CTCCTTTTGT AAAGCTGTAC	3720

776

CTCATTGAC	GACGGAGTAT	AAGCGCTTAG	CGCAAGCGCA	TGGTCTGAAT	TTAAAACAGG	3780
CTAAACCAAT	CACCATGGGT	ATGTGGATAG	GTGGTGACCG	TGATGGAAAT	CCATTTGTTA	3840
CAGCAAAGAC	CTTGAAGCAG	TCTGCACTCA	CTCAGTGTGA	AGTCATCATG	AACTACTATG	3900
ATAAAAAGAT	TTACCAACTT	TATCGTGAAT	TTTCTCTTTC	AACTAGCATT	GTCAACGTCA	3960
GCAAGCAAGT	CAGAGAAATG	GCTCGTCAAT	CCAAGGATAA	CTCGATTAC	CGCGAAAAAG	4020
AGCTTTACCG	TCGTGCCTTG	TTTGATATTC	AATCAAAAAAT	TCAGGCAACT	AAAACCTATC	4080
TGATTGAGGA	TGAAGAAGTT	GGGACTCGTT	ATGAAACCGC	CAATGATTTT	TACAAGGATT	4140
TGATTGCCAT	TCGAGATTCT	CTACTAGAAA	ATAAGGGCGA	GTCTTGATT	TCAGGTGATT	4200
TTGTGGAATT	ATTGCAGGCA	GTAGAGATAT	TTGGTTTTTA	CTTAGCATCA	ATTGATATGC	4260
GACAAGACTC	TAGCGTCTAT	GAAGCCTGTG	TGGCAGAACT	CTTGAAATCA	GCAGGAATTC	4320
ATTCTCGTTA	TAGCGAGTTG	AGCGAAGAAG	AAAAGTGTGA	CCTTCTCTTG	AAAGAATTAG	4380
AAGAAGATCC	CCGAATTCCT	TCTGCGACTC	ACGCAGAAAA	ATCAGAATTA	TTAGCAAAAG	4440
AATTAGCTAT	TTTTAAGACG	GCTCGTGTTT	TGAAAGATAA	GTGGGAGAT	GATGTCATCC	4500
GTCAGACCAT	CATTTACAT	GCAACCAGCC	TTTCTGATAT	GCTAGAATTA	GCTATTCTGT	4560
TAAAAGAAGT	AGGACTGGTG	GATACGGAAA	GGGCGCGTGT	TCAGATTGTT	CCCCTTTTTG	4620
AAACAATTGA	AGACTTGGAT	CATTCAGAGG	AAACAATGAG	AAAATATCTT	TCTCTTAGCC	4680
TTGCCAAAAA	ATGGATTGAC	TCACGAAATA	ACTACCAAGA	AATCATGCTT	GGCTACTCTG	4740
ACAGTAATAA	AGATGGCGGT	TACTTGTCAT	CATGTTGGAC	CCTCTACAAG	GCTCAACAAC	4800
AATTGACTGC	TATTGGAGAT	GAATTTGGCG	TTAAGGTTAC	CTTCTTCCAT	GGTCGTGGTG	4860
GTA CTGTCGG	TCGTGGTGGT	GGGCCAACCT	ATGAAGCCAT	TACATCTCAA	CCGCTCAAGT	4920
CTATCAAGGA	TCGTATCCGC	TTGACGGAGC	AGGGTGAAGT	AATTGGGAAT	AAATACGGTA	4980
ACAAAGACGC	CGCTTACTAT	AACCTTGAAA	TGCTAGTATC	GGCAGCTATT	AACCGTATGA	5040
TTACTCAGAA	GAAGAGCGAT	ACCAATACCC	CAAATCGTTA	TGAAACCATT	ATGGATCAAG	5100
TAGTGGACCG	TAGTTACGAT	ATCTACCGTG	ATTTGGTCTT	TGGTAATGAG	CATTTCTATG	5160
ATTATTTCTT	CGAGTCAAGT	CCAATCAAGG	CTATTTCAAG	TTTTAATATT	GGTTCTCGTC	5220
CAGCCGCTCG	TAAGACTATT	ACTGAAATCG	GTGGTTTGGC	TGCCATCCCT	TGGGTATTCT	5280
CATGGTCACA	GAGTCGTGTT	ATGTTCCCTG	GATGGTACGG	GTTTGGTTCA	AGCTTCAAGG	5340
AATTTATCAA	TAAAAATCCA	GAGAATATTG	CTATCTTACG	AGATATGTAC	CAAAATTGGC	5400
CTTCTTCCA	ATCGCTTCTT	TCAAATGTTG	ATATGGTTTT	GTCAAAATCA	AATATGAATA	5460
TTGCTTTTGA	ATATGCTAAA	CTTTGTGAAG	ACGACCAAGT	TAAGGCCATC	TATGAGACTA	5520



777

TTTTAAATGA ATGGCAAGTT ACTAAGAACG TTATCTTGGC TATTGAAGGA CATGACGAAC	5580
TCTTAGCTGA CAATCCATAT CTAAAAGCTA GTCTGGATTA CCGTATGCCT TACTTTAATA	5640
TTCTCAACTA TATTCAAGTTG GAGTTGATTA AACGCCAACG TCGTGGAGAA TTGTCCAGTG	5700
ATCAAGAACG ATTGATTCAT ATCACCATCA ACGGAATTGC GACAGGATTG CGTAATTCAG	5760
GTTGATAATT TTCAACAGTG AATGCTAAAA GTGAATATCA AAAAAATTCT AATAGACTAT	5820
<hr/>	
TGACAAGTAG TTTAAAAATG ATATAATTTA ACCATTGAGA AAAGTAATCA TACAACTTT	5880
TTAGAGAGTC TGTGGTAGCT GAAAACAGAT AAGTGGCAAT GATGAAAATT GGGCTGAATG	5940
CTATTTAGAA TTTGAAATTA TAAAAATTCG GTAAGCACAC CTTACAGTGC ATCTCGTTAT	6000
TGCGAGACTG AGCGATAGGG AAATTCCTTA TAATTGAGGT GGTACCGCGC ATCGACGTCC	6060
TCACACAAGT TTTTGTGTG AGGATTTTTT TGATGGAGGT TAGTATGGAA AGAAAACGAT	6120
GGCGTCGCTT GTTTAGATAA GTGAAATATG TTAAAGGAAA TAAAAAGGAG AAACAGAATG	6180
AAAAATAAAC GTTTAATTGG AATTATTGCT GCATTAGCAG TCTTAGTAGC AGGAAGCTTG	6240
ATTTATTCTT CAATGAATAA ATCAGAAGCT CAGAATAATA AGGATGAGAA GAAAAAACC	6300
AAGATTGGTG TGCTTCAATT TGTGAGCCAT CCATCCCTTG ATTTGATTTA TAAAGGGATC	6360
CAAGATGGAC TTGCAGAAGA AGGATATAAA GATGATCAAG TTAAAATTGA TTTTATGAAC	6420
TCAGAAGGTG ACCAAAGTAA GGTGCGACA ATGAGTAAAC AATTGGTTGC AAATGGGAAT	6480
GACCTTGTGG TTGGTATCGC AACACCAGCA GCCCAAGGGT TGGCTAGTGC AACAAAAGAC	6540
CTACCGGTTA TCATGGCCGC TATTACAGAC CCAATTGGTG CTAAGTTGGT TAAAGATTTG	6600
AAAAAACCAG GTGGCAACGT TACAGGGGTA TCTGACCACA ATCCAGCTCA ACAACAAGTT	6660
GAAGTATCA AGGCTCTGAC ACCGAATGTG AAAACAATCG GAGCTCTTTA CTCAAGTAGC	6720
GAAGACAATT CAAAA	6735

## (2) INFORMATION FOR SEQ ID NO: 105:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6516 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

CTAGAGGATC CCAGCAGGTA AATTGGCTTC AGCTGGCAAA AAAGTTGCCC TCGTTGAACG	60
CAGCAAGGCT ATGTACGGTG GAACTTGTAT CAACATTGGT TGTATCCCAA CTAAAACCTT	120

778

GCTAGTTGCT GCTGAAAAGG ACTTGTCTTT TGAAGAAGTC ATTGCTACTA AAAACACGAT	180
CACTGGTCCG CTCAACGGTA AAAACTATGC GACTGTTGCT GGTACAGGCG TAGATATCTT	240
TGATGCGGAA GCTCACTTCC TTTCAAATAA AGTCATCGAA ATCCAAGCTG GTGATGAAAA	300
GAAAGAACTG ACTGCTGAAA CAATCGTCAT CAACACTGGT GCTGTTTCAA ACGTCTTGCC	360
AATCCCTGGA CTGCTACAA GCAAAAACAT CTTTGACTCA ACAGGTATCC AAAGCTTGGA	420
CAAATTACCT GAAAACTTG GAATCCTTGG TGGCGGAAAT ATCGGTCTTG AATTTGCCGG	480
CCTTTACAAC AAACCTGGAA GCAAGGTCAC AGTCCTAGAT GCCTTGGATA CATTCCTACC	540
TCGTGCAGAA CCTTCCATCG CAGCTCTTGC TAAACAATAC ATGGAAGAAG ATGGCATTGA	600
ATTGCTTCAA AATATCCATA CTACTGAAAT CAAAAACGAT GGTGACCAAG TGCTTGTCGT	660
AACTGAAGAC GAAACTTACC GTTTCGACGC CCTTCTCTAC GCAACTGGAC GCAAAACAAA	720
TGTAGAACCA CTTCAACTTG AAAATACAGA TATTGAACTA ACTGAACGTG GTGCTATTAA	780
AGTAGACAAA CACTGTCAAA CAAACGTTCC TGGTGTCTTT GCAGTTGGAG ATGTCAACGG	840
TGGCCTTCAA TTTACTTACA TTTCACTTGA TGACTTCCGT GTTGTTTACA GCTACCTTGC	900
TGGAGATGGC AGCTATACAC TTGAAGACCG TCTCAATGTG CCAAATACTA TGTTTCATCAC	960
ACCTGCACTT TCACAAGTTG GTTTGACTGA AAGCCAAGCA GCTGATTGA AACTTCATA	1020
CGCTGTTAAG GAAATCCCCG TTGCAGCAAT GCCTCGTGGT CACGTAAATG GAGACCTTCG	1080
CGGTGCCTTC AAAGCTGTTG TCAATACTGA AACAAAAGAA ATTCTTGGAG CAAGCATCTT	1140
CTCAGAAGGT TCTCAAGAAA TCATCAACAT CATCACTGTT GCTATGGACA ACAAGATTCC	1200
TTCACTTAC TTCACAAAAC AAATCTTCAC TCACCCAACC TTGGCTGAGA ACTTGAATGA	1260
CTTGTTTGGC ATTTAAGTTG AGATTTAATC GTATCGAACA GCCCTCTTTG GGCTGTTTTT	1320
ACTTCTGCGG AATCTCAAAT CTGTCTTTCT CCTCTTTTAT GATATAATAG AAACATGAAC	1380
TTAAAACTA CTTTGGGCTT TCTTGCTGGG CGTTCTTCCC ACTTCGTTTT AAGCCGTCTT	1440
GGACCTGGAA GTACGCTCCC AGGGAAAGTC GCCCTTCAAT TTGATAAAGA TATTTTACAA	1500
AACCTAGCTA AGAACTACGA GATTGTCGTT GTCACTGGAA CAAATGGAAA AACCTGACA	1560
ACTGCCCTCA CTGTCGGCAT TTTAAAAGAG GTTTATGGTC AAGTTCTAAC CAACCCAAGC	1620
GGTGCCAACA TGATTACAGG GATTGCAACA ACCTTCCTAA CAGCCAAATC TTCTAAAACT	1680
GGGAAAAATA TTGCCGTCCT CGAAATTGAC GAAGCCAGTC TATCTCGTAT CTGTGACTAT	1740
ATCCAGCCTA GTCTTTTGT CATTACTAAT ATCTTCCTG ACCAGATGGA CCGTTTCGGT	1800
GAAATCTATA CTACCTATAA CATGATATTG GATGCCATTC GGAAAGTTCC AACTGCTACT	1860
GTTCTCCTTA ACGGAGACAG TCCACTTTTC TACAAGCCAA CTATTCCAAA CCCTATAGAG	1920

779

TATTTTGGTT TTGACTTGGA AAAGGGACCA GCCCAACTGG CTCACTACAA TACCGAAGGG	1980
ATTCTCTGTC CTGACTGCCA AGGCATCCTC AAATATGAGC ATAATACCTA TGCAAACCTG	2040
GGTGCCATA TCTGTGAAGG TTGTGGATGT AAACGTCTTG ATCTCGACTA TCGTTTGACA	2100
AAACTGGTTG AGTTGACCAA CAATCGCTCT CGCTTTGTCA TAGACGGCCA AGAATACGGT	2160
ATCCAAATCG GCGGGCTCTA TAATATCTAT AACGCCCTAG CTGCTGTGGC CATCGCCCGT	2220
TTCCTAGGTG CCGATTGCA ACTCATCAA CAGGGATTG ACAAGAGCCG TGCTGTCTTT	2280
GGACGCCAAG AAACCTTTCA TATCGGTGAC AAGGAATGTA CCCTTGTCTT GATTAAAAAT	2340
CCAGTCGGTG CAACCCAAGC TATCGAAATG ATCAAACCTAG CACCTTATCC ATTAGCCTA	2400
TCTGTCTCC TTAATGCCAA CTATGCAGAT GGAATTGACA CTAGCTGGAT CTGGGATGCA	2460
GACTTTGAAC AAATCACTGA CATGGACATT CCTGAAATCA ACGCTGGCGG TGTCGTCTAT	2520
TCTGAAATCG CTCGTGCCT CCGAGTGACT GGCTATCCAG CTGAGAAAAT CACTGAAACG	2580
AGTAATCTGG AGCAAGTTCT CAAGACCATT GAGAATCAAG ACTGCAAGCA TGCCTATATT	2640
CTGGCAACTT ATACTGCCAT GCTGGAATTT CGTGAACTGC TGGCTAGTCG TCAGATTGTT	2700
AGAAAGGAGA TGAACCTAATG GTTTATACTT CACTTTCCTC AAAAGATGGC AATTACCCCT	2760
ATCAGCTCAA CATTGCCCAC CTCTACGGAA ATCTCATGAA TACACGGGG ACAATGGAAA	2820
CATCCTCATG CTCAAGTATG TGGCTGAAAA ACTGGGAGCC CATGTGACCG TTGACATCGT	2880
TTCTCTCCAT GATGACTTTG ATGAAAATCA CTACGACATC GCCTTTTTCG GTGGTGGTCA	2940
AGACTTTGAA CAAAGTATCA TTGCAGACGA CCTACCTGCT AAAAAAGAGA GCATTGACAA	3000
CTACATCCAA AACGACGGTG TAGTTCTGGC TATCTGCGGT GGTTCCTAAC TATTGGGTCA	3060
ATATTATGTT GAAGCTTCAG GAAAACGTAT CGAAGGGCTA GGGGTCATGG GACACTACAC	3120
GCTCAACCAG ACCAATAACC GTTTTATCGG TGACATCAAG ATTCACAATG AAGATTTTCA	3180
TGAAACCTAC TATGGATTG AAAATCACCA AGGTCGTACC TTCCTCTCTG ATGACCAAAA	3240
ACCGCTGGGA CAGGTTGTCT ATGGAAATGG AAACAACGAA GAAAAGGTCG GTGAAGGGGT	3300
TCATTATAAG AATGTCTTTG GTTCCTACTT CCACGGGCT ATCTCTCTC GTAATGCCAA	3360
TCTGGCTTAT CGCCTAGTTA CTACTGCCCT CAAGAAGAAA TATGGTCAGG ACATCCAACT	3420
CCCTGCCTAT GAGGACATTC TCAGCCAAGA AATCGCTGAA GAGTACAGTG ACGTCAAAAG	3480
CAAGGCTGAC TTTTCTTAAA CAAAGGAAAA TGATATCAAA GAACTCCGTT ATCTGTGCGG	3540
AGTTTTTTGT CTTTTCTTTT ACCCTTCTCC CTGTCATTTT CTCTCATTTT TTGCCAAAAT	3600
AGAGGGGTAG AAAGAAGGTA GCATATGTCT AAATTACAAC AAATCCTAAC ATATCTTGAA	3660

780

TCAGAAAAAC TAGACGTCGC TGTGCTATCT GACCCCGTCA CAATCAATTA CCTCACTGGT	3720
TTTACAGTG ATCCCATGA ACGCCAAATG TTCCTCTTTG TCCTAGCAGA TCAGGAACCT	3780
CTCCTCTTTG TCCCAGCTCT TGAAGTAGAA CGTGCAAGTA GCACCGTTTC CTTCCCAGTA	3840
GTGGGCTATG TCGATTCTGA AAATCCATGG CAAAAAATCA AACATGCTCT TCCACAACCT	3900
GACTTCAAAC GTGTCGCTGT TGAGTTTGAC AATCTCATCT TGACCAAATA CCATGGTTTG	3960
AAAACAGTTT TTGAGACTGC TGAGTTTGAC AACCTCACTC CTCGTATCCA ACGCATGCGC	4020
CTCATCAAAT CAGCTGATGA AGTGCAAAAA ATGATGGTTG CAGGTCTTTA TGCTGACAAG	4080
GCTGTTTCATG TTGGTTTTGA CAATATTTCT CTTGATAAGA CTGAGACAGA TATCATCGCA	4140
CAAATCGACT TTGCCATGAA ACGTGAAGGT TATGAAATGA GCTTTGATAC CATGGTCTTG	4200
ACTGGTGATA ATGCTGCGAA TCCACACGGC ATTCCAGCAG CTAATAAGGT TGAAATGAT	4260
GCTCTTCTCC TCTTTGACCT GGGTGTTCTG GTCAATGGCT ATGCGTCAGA TATGACTCGT	4320
ACAGTCGCTG TCGGCAAACC AGACCAATTC AAGAAAGATA TTTACAACCT GACTCTTGAA	4380
GCCCAACAAG CTGCTCTTGA CTTTATCAAG CCAGGTGTGA CTGCTCATGA AGTGGACCGC	4440
GCTGCCCGTG AGGTCACTGA AAAAGCTGGT TATGGTGAGT ACTTCAACCA CCGTCTCGGG	4500
CATGGTATCG GTATGGATGT CCATGAATTC CCATCTATCA TGAAGGAAA CGACATGGTC	4560
ATCGAAGAAG GCATGTGCTT CTCTGTTGAA CCAGGTATCT ATATCCCTGG TAAAGTCGGT	4620
GTTCGTATTG AAGACTGCGG TGTGTTTACC AAGGATGGCT TCAACCTCTT TACAAGCACC	4680
AGCAAAGATT TGCTTTATTT TGATTAACT ATATAGCCCC TATGCTTTCC TTTCAAATA	4740
TCTAGGGGCT ATTTTATTGT CATTTTCTG CTATTATGCT AAAGAAATTG GCTGCAATAA	4800
TCTAACCCCTA AGTGTCTGGA ATGATAACGA GGGTGCTCTC CGCTTTTATC AAAGACAAGG	4860
GATGAAACCC CAAGAAACAA CAATGGAAAT GATAATTGAT TAAGAAGTCA TCTATCAAAA	4920
GATGTTAGAA AAAGTTCAAT TTCACTAGAA AATGAGGAAA ATCTCCCCAC AATAAAACGC	4980
ATAGTATCAG GTATTGTGTA CTGACCCCAA ACAGTTAGAC AATTAATTTA TCCGAAGGAT	5040
TTAGTTCTGT ACTGCACAGG ACTAAGTCCT TTTAGTTTTA CCTTAATTCG TTTGTTGTTG	5100
TAGTAATCAA TATAGTCTAT AATGACTTGT TCCAATTGGT TAAGTGATTT AAATGTTTTC	5160
TCATAGCCAT AAAACATTTT GGATTTTAAA ATGCCAAAGA AAGATTCCAT CATACCGTTG	5220
TCTTGGCTGT TTCCCTGCG TGACATAGAT GCTTGAATTC CCTTATCTC TAGGAACCGA	5280
TGATAAGAAT CGTGTTGGTA TTGCCAGCCT TGGTCACTAT GGAGAATCGT ATTCTCGTAG	5340
TGCTTCTCTT TGAATGCCTG TTCCAACATT GTTTGTAATT ATTCTAAATT AGGCGAACAA	5400
GAAAGATTAA AAGCAATAAT TTCGCTGTTA AAGCCATCTA AAAGTGGTGA TAAGTAAAGC	5460

781

TTTTGAGTAC TTGCTGGAAT GGCAAATTCA GTCACATCTG TGTAGCACTT TTCCATTGTT 5520  
 TTAGAGCCTT CAAATTGGGC TTGAATGAGA TTCTCTGCCT TCTTACCAAC GTCTCCTTTA 5580  
 TGAGAAGAAT ATTTTCGTTT CTTCGCATT TTAGCTTGTA AATTGAGTAC TTTCATCAAG 5640  
 CCTTGAAC TC TTTTATGATT TACCAGATAA CCACGATTTC TTAGTTCTAA ATGAACCCGG 5700  
 CGATAAGCAT AATTTCCTT GTGTTGATA AAGATGGATT GAATTTCACT TTTAAGCTCT 5760  
 TGGTCTTTAT CTGTTTGTG TAGCTGTTTC AAGTGATAGT AGTAGGTCCA ACGAGCTAGT 5820  
 TTAATGGCTT CTAGAAGAAG ATCTAACGAA AACTCAGTCA TTAATCTTG AACAAATTCT 5880  
 GTCTTTCTTC TTTCTCTTT TCCTCCTTCA ATCGGAGTTC TCTTAACCTT TTTAGGATGG 5940  
 CATCTCCGC TCTCAGGTAC TCTCCCTCTT GTTTTCTCAA CAATAGTATA CCCGTTTTTC 6000  
 CTGTATTGTG CTAGCCAGTT AAGAAGTATC GTACGACTTG GGAGACCGTA TTCAAGAGAA 6060  
 ACTCTATCTT TAGTCCAGCC TTCATGTCAG ACTTTATTAA CCCCAATTAT TCACCCCAA 6120  
 TCTAAAAACC ATCCAGAATC CTGCGCTTAG CTTAGATCCT GGATGGTTTC TTTTTTCACC 6180  
 CAATGGGTGT TTTTACTAG ACAAAAAAGA GTTTCCCTT TATGGTATAA GTGTAGAAAA 6240  
 AAACACAAAA AGAAAGGAAA CTCACATGAA CAGTTTACCA AATCATCACT TCCAAAACAA 6300  
 GTCTTTTAC CAACTATCTT TCGATGGAGG TCATTTAACC CAGTATGGTG GTCTTATCTT 6360  
 TTTTCAGGAA CTTTTTCCC AGTTGAACT AAAAGAGCGG ATTTCTAAGT ATTTAGTAAC 6420  
 GAATGAMCAA CGCCGCTACT GTCGTTATTC GGATTCAGAT ATCCWTGTCC AGTTCCTCTT 6480  
 TCAACTGTTA ACAGGTTATG GAACGGAATA TGCTTG 6516

## (2) INFORMATION FOR SEQ ID NO: 106:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 14654 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

TTTTCAACCC ATATCGTGGC TCCTGAATAC TACTTACTGA CAACTATGCT ATCAGAGACT 60  
 TCTCTACTTG TTTTCTATAT CATTTCATC CATAGAAAAC AACTCATCCA CTGGGACAT 120  
 ATCTTTAGCT ATACTGTTTC ATACTCTCTC TTTTCACTT CTTTGTAGC AATTATTTC 180  
 CTGATTAATT TCGTGTATCC TGATAGATAG GTCATTAAAT TGCCATTTTT GATTAATACT 240  
 GGTMTGATTG TCTTGCTATC AGCTATCTCT TATATTAGTC TACTTGTCTT CACAAAAGAT 300

782

AGCATTCTCT ATGAATTTT AAACCATGTC CTAGCCTTAA AAAATAAATT TAAAAAATCA	360
TAGGAGTTTA AAATGAAACA ACTAACCCTT GAAGATGCCA AACAAATTGA ATTAGAAATT	420
TTGGATTATA TTGATACTCT CTGTAAAAAG CACAATATCA ACTATATTAT TAACTACGGT	480
ACTCTGATTG GGGCGGTTTC ACATGAGGGC TTTATCCCTT GGGACGACGA TATTGATCTG	540
TCCATGCCTA GAGAAGACTA CCAACGATTT ATTAACATTT TTCAAAAGGA AAAAAGCAAG	600
TATAAGCTCC TATCCTTAGA AACTGATAAG AACTACTTTA ACAACTTTAT CAAGATAACC	660
GACAGTACGA CTAAATTTAT TGATACTCGA AATACAAAA CCTATGAGTC TGGTATCTTT	720
ATCGATATTT TCCCTATAGA TCGCTTTGAT GATCCTAAGG TCATTGATAC TTGTTATAAA	780
CTGGAAAGCT TCAAACTGCT GTCTTTCAGT AAACATAAAA ATATTGTCTA TAAGGATAGC	840
CTTTTAAAG ATTGGATACG AACAGCCTTC TGGTTACTCC TTCGACCGGT TTCTCCTCGT	900
TATTTTGCAA ATAAATCGA GAAAGAAATT CAAAAATATA GTCGTGAAAA TGGGCAATAT	960
ATGGCTTTTA TCCCTTCAAA ATTTAAGGAA AAGGAAGTCT TCCCAAGTGG TACCTTTGAT	1020
AAAACAATCG ATTTACCCTT TGAGAAATTA AGCCTTCCTG CACCTGAAAA ATTTGATACT	1080
ATTTTGACAC AATTTTATGG AGATTATATG ACCCTACCAC CAGAAGAAAA ACGCTTCTAC	1140
AGTCATGAAT TTCACGCTTA TAAATTGGAG GATTAGGATG CAATATTTAG AAAAAAAGA	1200
AATTAAAGAA ATTCAACTAG CCCTGCTGGA CTATATTGAT GAGACTTGTA AGAAACATGA	1260
TATTCCTTAT TTTCTCAGTT ATGGAACCAT GCTTGGAGCC ATCCGCCACA AAGGTATGAT	1320
TCCTTGGGAT GATGATATTG ATATTTCCCT TTATCGTGAG GATTATGAGC GTTTACTGAA	1380
GATTATTGAA GAAGAAAATC ACCCTCGCTA CAAGGTCTTT TCCTACGATA CATCTTCTTG	1440
GTA CTTCAT AATTTTCGAT CGATTTTGGA CACTTCTACT GTTATAGAAG ACCATGTTAA	1500
GTACAAGCGT CATGATACCA GCCTTTTCAT CGATGTCTTC CCAATTGATC GATTACAGA	1560
CTTGAGCATT GTCGACAAGA GCTATAAGTA TGTGGCTCTT CGTCAACTAG CTTATATCAA	1620
AAAATCACGA GCAGTTCACG GTGATAGCAA ACTAAAAGAT TTTCTTAGAT TATGTAGCTG	1680
GTACGCTCTC CGATTTGTCA ATCCTCGCTA CTTTACAAG AAAATTGATC AACTAGTCAA	1740
AAATGCTGTA ACCAACACTC CTCAATATGA AGGAGGAGTT GGGATCGGTA AGGAAGGGAT	1800
GAAAGAAATC TTCCAGTTG ATACCTTTAA AGAACTGATT TTAAGTGAAT TTGAGGGCCG	1860
TATGTTGCCT GTTCCCAAAA AATATGACCA ATTTTAAACC CAGATGTATG GCGATTATAT	1920
GACACCACCA TCAAAAGAAA TGCAAGAGTG GTATAGTCAT AGCATTAAAG CTTATCGCAA	1980
AAACTGATTG AGGGGGATTA TACAACTAC TAAGATAGAG GTTATTCAAA AACATAATTT	2040
TAGTAGAAAA TGAATACAT ATTCACCAA TAAACGCAT CATATCAAGG TTTTGA AAA	2100

783

ACCTTGATAT GATGCGTTTT ATAATTTTAA AGACTTTTTT CTATAGTAGA TTGAAATAAG	2160
ATGCGAACAA ATCAATTAGA AAATTCAAAT TAATTTATAG AAATATTTTA GTATTCCTGT	2220
GTA CTGTTCT AAATTCAGTC TGCTATATCT TATTTTTCTA TTTAAATCCG TCTGTAACA	2280
AAGCTACGAC TTTCAAGTAC CTTAAGCATG GCATTAGCTG TATCTAGCGC TGTGAAGAGG	2340
GGCACCCCGT GTTCAATGGC TGAACGACGA ATTTGCTCAC CATCTTCGTC AGCAGTTCGT	2400
TTTGTTCCCTA CTGTGTTAAT GATAGCTTGA ATTCTTCCTT TGCGTACAAA ACTTGGGATA	2460
TCCTTATCGT CATCACCAAT CTTACCAACA GGTGGGGCTT GCAAGCCATG ACTAGCAAAG	2520
AAGGCTGCTG TCCCTTCTGT CGCAAGGATT CCATAACCAA TGTTTTGGAA ACGACGAGCC	2580
AAGTTCAAGG CTTCTTCTTT GGCATCATCA GCGATGGTAA AGACGACATT ACCAAAAGTT	2640
GGCAAGTGTA GATAAGAAGC TTCAAAGGCT TTATAGAGAG CTTTTTCCAA AGTAGCATCA	2700
GAACCCATAA CTTACACCTGT TGACTTCATT TCAGGACCGA GCAAGCTGTC TACCTTAGCT	2760
AGTTTGGTAA AGGAGAAGAC AGGTGCCTTG ATATGAACAC GGGTGCTTTC AGGGTAAAGT	2820
CCATTTTGGT AGCCAAGTTC TGATAAACTT TGACCAAGAA TGAGTTTGGT CGCTACTTGA	2880
GCCATAGGAA TATTGGTTAC CTTAGATAGG AATGGAACAG TACGGCTGGC ACGTGGATTG	2940
ACCTCAATAA CGTAGACTTT TTCACTCTTG ATAACAACT GGATGTTTAT CATTCCAAGG	3000
CAGTGAAGAC CGATTGCTAA GCGTTTGGTG TAGTCTGCGA TGGTCTCCTG AACCTTTTGC	3060
GACAAGGTTT GTGGTGGGTA AACAGCCATT GAGTCACCTG AGTGGACACC AGCAGTTTCG	3120
ATATGCTCCA TGATACCAGG AATGAGTACA TTTTACCAT CTGAAATGGC ATCAACTTCG	3180
CACTCTTGCC CAACGATATA AGAGTCGACA AGAACTGGGT GGTCTGGACT AGCCTTAACA	3240
GCAGTTCGCA TGTAAGAACG AAGGTCTTCT TCGTTTTCAA CGATTTCCAT GGCACGTCCA	3300
CCAAGTACAT AAGATGGGCG GACAAGAACT GGAAGCCAA TCTTGCGAGC TGCAAGAGCT	3360
GCTTCTTCTT CATTGGTAGC CGTTTGTCTT GGTGGCTGTG GAATATCCAA TTCTTTGAGA	3420
GCTTGCTCGA AGAGGTCACG GTCTTCGGCA CGATCTAGGT CAGCAACCTG TGTACCAAGG	3480
ATGGTCACAC CTGCTTTTGC CAATGGCTCC GCAAGGTTGA TGGCTGTTTG ACCACCGAAC	3540
TGAACGATAA CTCCCTTTGG TTGTTCCAAG TCAATGACGT TCATAACATC TTGGAATGTC	3600
AATGGCTCAA AGTAAAGCTT ATCTGATACA GAGAAGTCTG TTGAAACGGT CTCTGGGTTT	3660
GAGTTCATGA TGATAGCTTC ATAACCAGCT GCCTGGATAG CCTTAACAGA GTGAACGGTT	3720
GCGTAGTCAA ACTCAACCCC TTGACCGATA CGGATTGGAC CTGAACCTAG GACAAGTACA	3780
GATTCTTTAT CAGATCTGAT AGATTCAATT TCCCAACCAT AGGTGAATA GAAATATGGC	3840

784

GTTTCGGAGT CGAACTCTGC CGCACAAGTG TCTACCATCT TATAAACTGG AACAACTCTTG	3900
TTTTCCAAGC GAAGTTGGCG AACTTTATCA TCAGTCGTTT CCCAGAGTTC AGCAATCTTA	3960
CGGTCTGAAA AACCATTAAAG TTTGGCTGTT TTCAAACTT CTAAATCTTG TGGATGAGCA	4020
CCCAATTCTT GCTCAATTTT AAAGATATGC AAGAGTTTAT CAAGATAGAA GATATCAATT	4080
TTTGTAAGCT CTGCAATTTT TTCAGGTGTG TAGCCACGAC GAATGGCTTC TGATACGTAG	4140
AAGAGACGGT CATCTTGGGC TTTGACAACC TTTTCAATCA AGGCATCATC AGAACTGCT	4200
GCAAGTTCAG GTATTTCAAT GTGGTGCACC CCAATTTCAA GGGAGCGGCA GGCCTTGAGA	4260
AGAGATTCCT CGATGTTACG ACCGATTGCC ATGACTTCTC CAGTCGCCTT CATTTGTGTA	4320
CCGAGACGGC GTTACCCTT TTCAAACTTG TCAAATGGGA AACGTGGAAT CTTAGCAACT	4380
ACGTAGTCAA GGGCTGGTTC AAACATGGCA TAGGTTGAAC CTGTAAGTGG GTTTATAACC	4440
TCATCCAAGG TCAAACCTAC TGCAATCTTG GCAGCCAACT TAGCAATCGG ATATCCTGTC	4500
GCTTTAGAAG CAAGGGCTGA CGAACGTGAT ACACGAGGT TTAATTCGAT AACATAATAC	4560
TTGAAGCTGT TAGGATCAAG AGCTAGCTGA ACATTACATC CACCTTCAAT CTTGAGGGCA	4620
CGAATAATGC TCAAGCTCGC ATCACGAAGC ATTTGGTTTT CATAGTCTGA CATGGTTTGC	4680
GCAGGGGCAA ATACAATGGA ATCCCTGTG TGAATCCAA CTGGGTCAA GTTTTCCATG	4740
TTACAAACAA CCAAGGCATT GTCAGCTGAG TCACGCATCA CTTCTGATTC AATTTCTTG	4800
AAACCGGCAA TCGAACGCTC AATCAAACAT TGGGTAACAG GTGACAATTT CAAACCATTT	4860
TCAGTGATTT CACGCAATTC TTTCTCGTTG GCACACATAC CACCACCAGT ACCACCAAGG	4920
GTAAGGCTG GACGAACGAT GACTGGGTAG CCAATTGTCG CTGCAAAGGC AACTGCTTCT	4980
TCTACTGTGT TAACAATTTT AGATTCTGGA ATGGGTGTT CAAGCTCTT CATCAATTGT	5040
TTAAAGAGGT CACGGTCCTC CGCTTGGTCA ATGGCAGATA ATTTGGTACC CAGAAGTTCA	5100
ACGCCAAGCT CGTCTAGGAT ACCATTTTTA GATAATTCCA TGGCCATGTT GAGACCTGTC	5160
TGACCACCGA GTGTTGGTAG CAAGGCATCT GGACCTTCTT TACGAAGAAT ACGTGTACAA	5220
AACTCAAGTG TAATCGGTTT AATGTAAACC TTGTCAGCAA TTCTCTGTC CGTCATGATG	5280
GTGTCAGGAT TTGAGTTAAC CAAAACAACC TCATAACCTT CCTCTTTCAA CGACAAGCAA	5340
GCCTGAGTCC CAGCGTAGTC AAACCTCAGCA GCCTGACCAA TAATAATCGG ACCAGAACCA	5400
ATCACCATAA TTTTTTGAAT ATCAGTACGT TTAGGCATAT ATAAGATATT AAGGGTGTCA	5460
AGCGGACAAA GCTAAAATAG GAGTTATGAC GAAGAACTGT CAGTTCTAGG AATAACTATC	5520
TTTTTAGCAC CGTCCGTAGC CCGTATTCAG TTCAGCAAAT ACGGAGCACC CTTCTCCTTT	5580
CTATTCGTCG CCTCTCAGGG CGACATTAAA TAAGATACAA AGGACGAATA GAAAGCGATT	5640



785

GAATTTTAGG AAATCAAGGA AGGATTGACA ATCCAAGTTG GTTCTCTAC ATTCTGAGCT	5700
TTCCGTCGGT GTTCAGTTAC ATAAATTCTC CGACGAGCTT TTA CTCTGTTT	5760
TGTTTAAAAA CTTCATCAT CTCGATAAAC TCGTCAAATA GGTAGCTAGC GTCGTGTGGC	5820
CCAGGAGCTG CATCTGGGTG GTATTGAACA GAGAAAGCAG GTTGGTATCT GTGGCGCACA	5880
CCTTCCACTG ACTTGTCATT GATTTCTTCG TGGGTAATAA TCAAGTGCTC TGGCAAATCC	5940
TCGCGGCTGA CTGCATAACC ATGGTTCTGG CTGGTGAAGT CTACTCGTCC TGTTGCGATT	6000
TCACGTACCG CATGGTTGAA TCCACGGTGG CCAAACCTCA TCTTATAGGT CTTAGCCCCG	6060
TTTGCCATTG CAAAGAGTTG GTGTCCCATTA CAAATACCAA AGATTGGAAT TTTTCCTTGT	6120
ACACCGCGAA TCATGTCGAG TGCTTGTTGA ACGTCTTCTG GGTACCTGG ACCATTTGAC	6180
AACATAAATC CGTCAGGATT GAGATGGAGA ATTTCTTCAG CCGTGTCTGA ATAAGGAACA	6240
ACTGTCACGT TACAGTTGCG TTTAGAAAGT TCACGTAGGA TTGAGTGCTT GAGACCAAAG	6300
TCCACTAGCA CCACGCTCAA ACCAACTCCT GGAGCTGGAT AAGACGTTTT AGTAGAAACC	6360
TGTTTGATAT TGTCTGTGCG TAAACTGTT GCTTGGAGCT GGTCCGTCAC ATGGTCCATA	6420
CTGTCCCCAA CATGGGTCAA GGTGTCACGC ATAGTACCAT GCTTACGGAT AATCTTGGTA	6480
AGAGCACGCG TATCAATTCC TGAAATCCCT GGAATTTTCT TGGCTTTCAA AAATTCATCC	6540
AAGGTCATTT GGTGCGCCA GTTGCTAGCT CTACGCGCTT CTTCAAAAAC AACGACTCCC	6600
TTACAAGTTG GAATAATGGA TTCATAATCA TCACGATTAA TACCATAATT TCCTACCAA	6660
GGATAAGTAA AGGTCAAGAT TTGTCCATTA TAAGACTGCT CTGTAATGGA TTCTTGGTAG	6720
CCGGTCATCC CTGTATTAAA GACGATTTCTG CCTGTTACAT CAATATCTGC TCCGAAGGCC	6780
TTGCCTTCAA AAATGTGCC ATCTTCTAAT ACTAGAATTC TTTTGTCTAT ATTTTCACCT	6840
CTCGTGGACG CTCACTGGCG TCTTTTAACG TCTTGTGTTT TAGTTGGCGT TTCTACTCGC	6900
TAGTACGGAT TCTAAGATTG CCATTCGAAC AAAGACACCA TTGGTCATTT GTTGACAAT	6960
CCGTGATTTT GGTGCTTCAA CCAAGTGGTC TGCTATTTCT ACATCACCAT TGATTGGAGC	7020
TGGGTGCATG AGGATTGCTG TTTCTTTCAA ACGATCGTAA CGTTCTTGAG TCAAGCCATG	7080
TTGGGCATGG TAGTCTTCTT TTGAAAATAC AGCTCCACTA TCATGGCGTT CGTGTTCAC	7140
ACGAGAAAC ATCATGACAT CAACCTGATC AATGATTTCA TCAATGGTTA CAAACTGTCC	7200
ATAGTCTGCA AACTCTTGAC TTCTCCATTC CTCAGGTCCA GCGAAAAAGA GTTCAGCTCC	7260
CAACCGTTTC AAAATCTGCA TATTGGATTT GGCAACGGT GAGTGGTCCA AGTCACCTGC	7320
AATAGCAACT TTAAGACCCT CAAAGTGGCC AAATTCCTCA TAAATGGTCA TCAATCAAG	7380

786

CAAGCTCTGG CTAGGGTGTT GGCCCGAACC ATCTCCACCA TTGATGATGG AAGTCGTAAT	7440
CGTTGGACTA GCAATCAATT CTCTATAGTA GTCGACCTCT GGATGGCGAA TCACACAGAC	7500
ATCCACTCCT AAAGCAGACA GAGTCAAAAT GGTGTCATAA AGTGTCTCAC CCTTATTAAC	7560
CGAGCTAGTC TTCACATCAA AGTCAAGTCG TTCCAATCCA AGTTTAATCT CTGCGACTTC	7620
AAAGGACTTA TGTGTCCGTG TAGAATCCTC AAAGAAGAGA TTGGAAACAA TCGGATGGTC	7680
TTCATAGGGA AGCTGGGCTC CATTTTAACTA CTCAATTCCT CGCTTGATCA ATTTCAATTAC	7740
TTGATCGACA GTGAGGTCTT CCATGGACAC CACATGGTTC AATGCTTGTT GATTTTCTGA	7800
CATGGCTACT CCTTTAACTT TCTAAGCTTC TTCAGTAATC AGAACTCTGT CTTGGTCATC	7860
AAGTTCGTG ATCTCTACGA TGATTTCTTC AGAACGACTG GTTGGGATAT TTTTCCAAAC	7920
GTAATCTGGA CGGATTGGCA ATTCTCTATG TCCACGATCG ACTAGAACTG CTAAACTCAC	7980
ACGCGCAGGA CGACCATGAC CGACAATATT ATCAATAGCA GCACGGATGG TACGACCTGT	8040
ATAGAGCACA TCATCCACCA AGATAACTTC GCGGTCTGTC ACATCGACAG AAACCAAAGA	8100
AGTATCTTCT CCACTTTTAA CATCATCAGC GAAAGGTTTA GTATCCAATT CCACAACAGG	8160
AACTGAAAGA TTTTCTAACT GCTTCAAACG TTCTTGATT CGGTGGGCAA TAAAGACACC	8220
ACGAGTTTAA ATACCAGCCA AGACGATCTT ATTCAAATCT TTGTTGCGTT CGATAATCTC	8280
ATAAGTAATA CGCGTAATCG CTCGTTTGAC GGTCAATTCG TCTACAACTT CTTTGTGTTT	8340
CATGACAAAC CTCCAAAAG AAAAGTCTCC TTAAACAAGG AGACTTGAAA TTTATAGCCA	8400
AGCGAGCCCT ACTGCACACA GTATAGACTT CACCCTTCTA CTTTATCGCG CTCCTGCTT	8460
GCCTCACGGG ACAGGTTTAA AGGAATATTT AGTTATCATT TACTATAGCA CAAAGCATGC	8520
TTAAATCAA GCAAAAAGTT TCAATGTAGC ATCTTACAAA TTGCTAAAAT CATATAATTG	8580
TGGGTACTGG TCACACTCTG GATTTTTTGG ATGGCAAATG GCTCTTCCAA AATAAATCAT	8640
GGCCTGATGG GCAGCTAACC ACTGCTCAGG CGGCAAGATA TCCATGACCC GCTTTTCCAC	8700
CTCAAGTGGC GTCGCTGATT TTTTGACAAT ATCGTGGTGT TTGCAAATAC GCTCCACAAG	8760
AGTATCCACT GCAAAGGCTG GAATTCACAA TCCTACACTC ATGACAACAT TGGCTGTCTT	8820
GGACCAACA CCTGCCAAAC TCTCCAATTC TTCACGTGTC TGAGGGACTT GACCATCAAA	8880
ATCGTCTAGT AACTGTTGGG CACATTTTTT AAGGAATTTA GCTTTATTCC GATACAATCC	8940
CAAGCGAGAA ATATGTGAAG CAATCTCACT CTCTGTGCT ACAGACATAG CTTGGGGTGT	9000
TGGAAAGGCA ACAAAGAGAC CTGGTGTTGG CTTATTTACC GCTGCATCTG TCGTCTGGGC	9060
TGATAACATG ACCGCAACCA GGAGTTCAAA ATGATTGGTA AAATCAAGAC TAGGCTTGGC	9120
ATCTGGGAAG AGGGCAATGA TTTCTTCTAG CACCTTTCCT GCTCGTTTTT TTGACAAGAC	9180

787

CATTATTCAT CTCCGTCAA TAGTCCTTGT AAGCCAGCAA AAGGACTGTT TTCTTCTTTC	9240
TTTACTGCTT TTTGAGCTTG GTATTCTTCC TCTGTCATGA TTTGCCAGTC ATTTCTGAG	9300
ATAAATCCTT GACCAGCTTC TTCTTCAGCC GTCAAGACCT TGATAGGAAT GTTTAGCAGG	9360
ATATTGTCTG ATACACTCTC AGCAAGGTCA AGCTCCCCAT TTTCGATGGG CAAGACCAAG	9420
TCATCATCTA AAACCTCTTG ATCTAGCTGG TTAGTTGCGC CTTCCATGAA AACTTCCGTG	9480
ACTGGATAAG ATTCAACTAA CTCAACTGGC TCCATACTGC GACTCGACGC AAGAACAATG	9540
GTATAAGATA GTTGATAATC TAAGAAATAC ATACGGTCTT CATATTGTAC TTTCCCAACT	9600
GCAAGGATAT CTTTACATC TAAAATTTCT TGATTACGTG CACGCAGGTC ATCAACTAAA	9660
TCTAACGTTT GTTCAAAGTT CAAACCTTCA GACTGCTTAC GAATTTCTTG AATATTTAAT	9720
TTCATACTTC CTCCATAAAG ATTTACTCTC TTGATTATAC CATGAAAAGG CTACAAATCA	9780
GCACACCAAA CTTTGTAATT AAAATTCAAA ATTTTAACAT ATTTACTATG ATAGTTTTAT	9840
TTTTTAGTGC TATACTATAG GGAAAGAGTA CATCAGATCA AGGAGGATGC TCACATGGAA	9900
GACAAGAAAC TCATTCAACT CCTATCCAAG TTAATAAAAA GCTACCAAAA CTGTAACAG	9960
GGTACGGCAG ATGATATTCG ACTACAAGAG CTGCTAAACA CTACTATGCA AGAGCTCAAA	10020
AAAACGGAAC AGTTGAACAA CAGTATCTTA ATTGATCTTG AGAAATTTTA CCAACCTACC	10080
AGTCTTCTGA TTGGACTGGG TAGCCTAAAA CTAAACGATC AAGCACGCAC TGCTTGGCGA	10140
AACTATGATA AATTCATTA CGATCATGTC AAACACGTAC TAAGTCTCTA TGGACCTGTT	10200
TTTGAATTTT AGAGCATAGA ATTTCCAGTT TTCTGTTGAC AAAATTTTCT TAAAGGTATA	10260
ATATAAGAT ACTAATACTC GGAGGTAAGG GAGACATGAA CAACTAAGTC TATCAAATAA	10320
AGAACCTTTA TTTAGTAGAT CTTGTTTTTG TCTCTTTTGG TGTGCTCTTT TATGCTCTTT	10380
TTCTGGCATG TTAATAGAGT TTTTTTGACA TAGACTTTGG GCTCTACTAG GTAAAGTAGA	10440
GCTTTTGTGTT ATGCACTATG AACATTCTAG AAAGGGAAAT CATATGATAA AAATCAATCA	10500
TCTAACCATC ACACAAAACA AAGATTTACG AGATCTTGTA TCTGACCTAA CCATGACCAT	10560
CCAAGACGGG GAAAAGGTTG CTATTATTGG TGAAGAAGGA AATGGCAAAT CAACCTTACT	10620
TAAAATTTTA ATGGGGGAAG CTTTGTCTGA TTTCACTATC AAGGGAAACA TCCAATCTGA	10680
CTATCAGTCA CTGGCTTACA TTCCTCAAAA AGTCCCTGAG GACCTAAAAA AGAAAACCTT	10740
ACACGACTAC TTCTTTTAG ATTCTATTGA TTTAGACTAC AGTATCCTCT ATCGTTTGGC	10800
GGAGGAATTG CATTTTGATA GCAATCGTTT CGCAAGTGAC CAAGAGATTG GCAATCTATC	10860
AGGGGGCGAA GCTTTGAAAA TTCAGCTTAT CCATGAGTTA GCCAAACCCT TTGAGATTCT	10920

788

ATTTTGTAGAT GAACCTTCAA ATGACCTAGA CCTTGAGACA GTTGATTGGC TAAAAGGCCA	10980
GATTCAAAAG ACCAGGCAAA CCGTTATTTT CATTTCCCAT GATGAAGACT TTCTTTCTGA	11040
AACGGCAGAC ACTATTGTTC ACTTGCGACT GGTCAAACAC CGTAAAGAAG CGGAAACGCT	11100
AGTAGAGCAT TTAGACTATG ATAGCTATAG TGAGCAGAGA AAGGCTAATT TTGCCAAACA	11160
AAGTCAGCAA GCTGCTAACA ACCAAAGAGC CTACGATAAA ACCATGGAAA AACATCGGAG	11220
AGTTAAGCAA AATGTAGAAA CTGCGCTTCG AGCTACCAAA GATAGTACTG CCGGTCGCCT	11280
ATTGGCTAAA AAGATGAAAA CTGTCCTCTC ACAAGAAAAA CGCTACGAAA AGGCAGCTCA	11340
GTCCATGACT CAAAAGCCAC TTGAAGAGGA ACAAATCCAA CTTTCTTTT CAGACATCCA	11400
ACCATTACCA GCTTCTAAAG TCTTAGTCCA ACTGGAAAAA GAAAATTTGT CCATTGACGA	11460
CCGAGTTTGT GTTCAAAAAC TACAACCTAAC TGTCCGTGGC CAAGAAAAAA TCGGTATTAT	11520
CGGGCCAAAT GGTGTTGGGA AATCAACTCT GTTAGCCAAG TTACAGAGAC TTCTGAATGA	11580
TAAAAGAGAG ATTTCACTTG GTTTTATGCC ACAAGATTAC CACAAAAAAC TGCAATTGGA	11640
TTTATCCCCA ATAGCCTATC TCAGTAAAAA TGGGGAAAAA GAGGAACTAC AGAAAATCCA	11700
ATCTCACCTA GCTAGTCTCA ATTTCACTTA TCCAGAAATG CAGCATCAAA TTCGCTCCTT	11760
ATCTGGCGGA CAACAGGGAA AACTCCTGCT TTTGGATTTA GTCCTGCGCA AACCAAACTT	11820
TCTCCTGCTG GATGAACCCA CACGAAACTT TTCTCCCACT TCTCAACCCC AAATCAGAAA	11880
ACTCTTTGCT ACCTATCCAG GCGGTCTCAT CACTGTTTCG CATGACCGTC GTTTCTTAAA	11940
AGAAGTCTGC TCGATCATCT ATCGCATGAC AGAACACGGT TTGAAGCTAG TTAATTTAGA	12000
AGATTTATAA ATTTGCAACA TAGCAAAAAT CCAGAGACGA CCTCTGGATT CTTTACATC	12060
TGTTTTAAAC GTTCAATCCG TTCTGAGATA GGTGGGTGGG TATAAAAGAG TTTTGGAAAC	12120
CCCCCACCTT TCTTAGGATC ATTGATATAA AGGGCACTGC TAGCATCATC GACGTGGCGA	12180
CTCATAGGTT TGCTATTGTC CAACTTATCT AGGGCATTAA TCATTCCCTG GGGATTGCGA	12240
GTCAGCTCGA CACTAGATGC ATCTGCCAGA AATTCCCTCT GACGAGAAAT AGCGAGCTGA	12300
ACCAAGGTTG CAGCGAGAGG TGCCAGTACA ATAGCTAGTA GGGAAACCAC TAGCATAATG	12360
ATTTCAAGAC CATTTCCATC TCGGTCATCA TCACTTCGTC TCGGACCTGC TCCACCCAC	12420
CACATCATAC GACCTGCCAT ACTAGAAAGC ATGGTGATAG CACTAGCAAG GGCAACTGCA	12480
ATAGTCGAAA TACGGATATC ATAATTACGA ATATGACTGA CTTTATGTCC CATAACAGCT	12540
TCTAGTTCTT CACGATTCAT GATAGCTAGT AGACCTGAAG TCGCAGCAAC AGCCGCATTT	12600
TGAGGATTAG AACCTGTCGC AAAGGCATTT AAGGCTGGAT CATCAATGAT GAAAACACGG	12660
GGCATAGGAA TCTGAGCGAC CAGAGCCATA TCTTCCAATA CATGGTAGAG GTCTGGTGCC	12720

789

GTTCGTCAT CCACCTCACC CGCTCCATTC ATGGACATGA CAATCTCTGT CGATTGAAAA	12780
ATCATAGACA AAGCGTAGAT AAAGCCGATA ATCAGTGCAA TAACCAAACC ACCAAGTCCA	12840
GATCTTATAA AGAGATAACC AACCCGATAA CCAACAAGAG CTAAGAGTAG GAAAAATACC	12900
AGCAACAAAA TCCAGGTTTT TCGTTTATTG CTTGCAATTT GATCAAACAA CATCTTAGTC	12960
ACCTAAACCG CTAAAATCAA CTTTAGGAAC CGACTTTTCC TCTTCAGGTG TTTGAAGGAA	13020
ATCTGCCGCT TTAAATCCAA ACATTCCAGC GATAATATTG CTCGGGAAAG TTTCTAATTT	13080
TACATTGTAG TTGCTGACAA CACTGTTATA GAGTTGACGA GAGTAAGAAA TTTTATTTTC	13140
TGTGTTTGTG AACTCCTCTT GCAATTTAAC AAAGTTAGCA CTAGCTTTCA AATCTGGATA	13200
GCTTTCTGCA ACTGCAAAAA TACCTGAAAC CTGACGAGTG AGGGCATCAC TGGCTTTTCAT	13260
AGCTTCTGCT GGTGAAGTCG CTGCCGCCAC TTGGTTACGT AGTTCTGCCA CCTTTTCAAG	13320
GGTAGAACCT TCATATTTGG CATAACCTTT TACAGTCTCA ATCAAGTTTG GCAAGAGGTC	13380
ATTGCGACGT TTCAACTGAA CATCAATCTG ACTCCAAGCC TCCTTGGTTT GCATACGATT	13440
TTTAACCAAA CCGTTATAGC TAACAATCAC AAAAATAACA ATAAGAGCGA TAACTCCAAG	13500
AATAATCCAA GTCATAATAT AAGTCCTTTC TGCTTTTAGA TTAGTACCAG TATATCAAAT	13560
TTTCTATGAT TGTGGTAAAA TAAGATGATA CTAAAGAAGG AAATAACTAT GAAACCAAAA	13620
ACATTTTACA ACTTGCTTGC CGAGCAGAAT CTTCCTTTT CGGACCAGCA AAAAGAACAA	13680
TTTGAACGTT ATTTTGAGCT CTTGGTCGAG TGGAATGAGA AGATTAATTT GACGGCGATT	13740
ACGGACAAGG AAGAAGTTTA TCTCAACAT TTTTACGATT CGATTGCACC CATTCTTCAA	13800
GGTTTGATTC CCAATGAAAC TATCAAACTT CTTGATATCG GGGCTGGGGC AGGATTTCTT	13860
AGTCTACCAA TGAAAATTCT CTATCCGGAG TTAGATGTGA CCATTATTGA TTTACTCAAT	13920
AAGCGCATCA ACTTCCTACA ACTCTTGGCT CAAGAAGTGG ATTTGAACGG AGTTTCATTTT	13980
TACCACGGAC GTGCCGAAGA TTTTGCCCAA GACAAGAACT TCCGTGCTCA ATATGATTTT	14040
GTAACAGCTC GTGCGGTTGC CCGTATGCAG GTCCTATCTG AATTGACTAT TCCCTACCTT	14100
AAGGTTGGTG GCAAACTATT AGCACTCAAG GCTAGCAATC CGCCTGAGGA ATTATTAGAA	14160
GCTAAGAAATG CCCTCAATCT CTTTTTTAGT AAGGTGGAAG ACAATCTCAG TACGCCCTAC	14220
CGAATAGAGA TCCGCGCTAT ATCACAGTGG TAGAAAAGAA AAAAGAAACA CCAAATAAAT	14280
ATCCACGTAA GGCTGGTATG CCAAATAAAC GCCCACTTTA AATTTTTTAG TAAACAAATG	14340
TTTACAAAAT CAGCCTCGCT CTTTTATTTC TAGGCTCGGG AAAAAATGAT TTACAAAATC	14400
AGCCTCGCTC TTTTATTTCT AGGCTCGGGA AAAAAATGATT TACAAAATCA TTTTTTCTG	14460

790

CTATACTATC CTAAGCAAAG GTTTTAAATG TCATCCCGTG AGGTGACGAA GACGCAGAAA	14520
TATTTAAAC TCTTTAAAT CTAATTTTA AAGAAGTCTT ACTCTGAGGG CCTATTGCTG	14580
TAAAATAATG GGCTCTTTT TGATGCCCAA AAGTGAGGTT TATATGAAAC AAGAATCAAC	14640
TGTTGATTG TTAC	14654

## (2) INFORMATION FOR SEQ ID NO: 107:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6405 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

AGAAAAATCT GCTTTACAGA AAATAAAAT AATAGGAGAA AATCTATGTC AGATTGAAA	60
AAATACGAAG GTGTCATTCC AGCCTTCTAC GCATGTTATG ATGATCAAGG AGAAGTAAGC	120
CCAGAACGTA CGCGTGCCTT GGTTCATAC TTCATTGATA AAGGTGTTCA AGGTCTTTAT	180
GTCAATGGTT CTTCTGGTGA ATGATCTAC CAAAGCGTTG AAGATCGCAA GTTGATTTTG	240
GAAGAAGTCA TGGCGGTAGC AAAGGTAAAT TGACCATTAT TGCCCATGTT GCTTGCAATA	300
ATACTAAAGA TAGTATGGAA CTTGCTCGCC ATGCTGAAAG CTTGGGAGTA GATGCTATTG	360
CAACGATTCC ACCAATTAT TTCCGCTTGC CAGAATACTC AGTTGCCAAA TACTGGAACG	420
ATATCAGTTC TGCAGCTCCA AACACAGACT ACGTGATTTA CAACATTCCT CAATTGGCAG	480
GGGTTGCTTT GACTCCAAGC CTTTACACAG AAATGTTGAA AAATCCTCGT GTTATCGGTG	540
TGAAGAACTC TTCTATGCCA GTTCAAGATA TCCAAACCTT TGTCAGCCTT GGTGGAGAAG	600
ACCATATCGT CTTAATGGT CCTGATGAGC AGTTCCTAGG AGGACGCCTC ATGGGGGCTA	660
GGGCTGGTAT CGGTGGTACT TATGGTGCTA TGCCAGAACT CTTCTTGAAA CTCAATCAGT	720
TGATTGCGGA TAAGGACCTA GAAACAGCGC GTGAATTGCA GTATGCTATC AACGCAATCA	780
TTGGTAACT CACTTCTGCT CATGGAAATA TGTACGGTGT CATCAAAGAA GTCTTGAAAA	840
TCAATGAAGG CTTGAATATT GGATCTGTTC GTTCACCATT GACACCAGTG ACTGAAGAAG	900
ATCGTCCAGT TGTAGAAGCG GCTGCTGCCT TGATTCTGTA AACCAAGGAG CGCTTCCTCT	960
AATCTAAAAG GAGGTATTTA TGACATATTA CGTTGCAATT GATATCGGTG GAACCAACAT	1020
CAAGTATGGT TTGGTTGATC AAGAGGGGCA ACTTCTTGAA TCGCATGAAA TGCCAACTGA	1080
GGCGCATAAG GGTGGACCTC ATATCTTACA AAAGACCAA GATATCGTAG CTAGTTATTT	1140
AGAAAAGGC CCAGTAGCAG GTGTTGCCAT ATCTTCTGCT GGGATGGTGG ATCCGGATAA	1200

791

GGGTGAGATT TTCTATGCTG GGCCGCAAAT CCCTAACTAC GCAGGCACCC AGTTCAAAAA	1250
GGAAATCGAA GAAAGCTTTA CTATTCCTTG TGAGATTGAA AATGATGTCA ACTGTGCAGG	1320
TCTTGCTGAG GCAGTATCTG GTTCAGGCAA GGGAGCAAGT GTGACACTTT GCTTGACCAT	1380
TGGAACCGGT ATCGGTGGTT GCTTGATTAT GGATAGGAAA GTCTTCCATG GTTTTAGCAA	1440
TTCAGCCTGT GAAGTCGGGT ATATGCATAT GCAGGATGGA GCTTTTCAAG ACTTGGCTTC	1500
TACAACAGCT TTAGTGAAAT ATGTAGCTGA AGCCCATGGA GAAGATGTTG ATCAGTGGAA	1560
TGGCCGTAGA ATTTTCAAAG AAGCCACTGA AGGAAACAAA ATCTGCATGG AAGGTATTGA	1620
CCGTATGGTT GACTATCTAG GAAAAGGTCT GGCAAATATT TGCTACGTTG CCAATCCAGA	1680
AGTGGTTATT CTTGGTGGTG GTATCATGGG GCAAGAGGCT ATCCTCAAAC CTAAGATCCG	1740
TACAGCCTTG AAAGAGGCTT TGGTACCAAG TTAGCAGAA AAAACACGAT TAGAATTGTC	1800
CCATCACCAA AATACAGCAG GGATGTTGGG TGCATATTAT CATTTTAAGA CAAAACAATC	1860
CTAGTTTGGC TCAGCCAAAC TAGGATTTTC TTACACGTTT TTGTCTACGA TAGCCGTTGA	1920
GTTTTTTATT TTCCCAGTAG CTATTAAAGA TTTTTTCCTT GCTTTCGCGA TTGATTTCCA	1980
AAAAGTAGGC ATAAATCAAA TCGATAAAGA AGAGCATAGG AAGTTGAGCG GATATTCGTT	2040
GGATATAGGA GGGTTGGCTG TGGGTGGCTA CAAGAACAGT CTCTGTATAG GTCTGGCTAT	2100
CTTTATTGGG AACACTTGTA AAGAGTACAG TCTTTGCCCC CATCTCCTTA GCATCTAATA	2160
GACTATCTAA AATAGAAGGA GTTGAGCCTG AAAGTGAGAA GCCCAGTACT AGACAATTTT	2220
CATCCATGAT GCTGGTTGTC CAGGCAAAGC CGTCTGGTC TGTCAAAGCT TCGCAGACCA	2280
CACCTAGTCG CATAAAACGT AATTTCATTT CACGGGCGAC GAGGCCAGAA CTCCTGTTC	2340
CAAAGAAGTA GATACGCTCA GCATCTTCGA TTAGCTGGGC AATTCGTTCT AGTTGGATTT	2400
CGTCAATCAA GTCTTGTGTT TGTTCCCTCA TATTGCTATA ACTTCTGAGG ACTCGTTTGG	2460
TCAGTGGACT GTGCTTGGAG ACTTGGTTGG CTTGATTTTC TGCCTGATGT TCGTATTGGA	2520
AAATAAATTC TCGGTAGCCA GTAAAGCCAC ACTTTTtagc AAAGCGGGTC AAAGCAGCTT	2580
GAGAAATATG TAATTTTGG GTGACTTGTT GAGAAGATAA ATCATCTGTA ATCGTTTCAG	2640
CTTGCAAAAA ATAGCGAGCG ATTTCTTGTT CTAGGTCTGT CATTTCTTCA AAATGTGAAT	2700
CAATGATAGT TCGGATATCT GGTTCGTCCA TAGGGAAAGC TCCTTTACAT GAGTCATACT	2760
GGAAGACTAG ATCAGAGAAT AGTCACACTT CATTATAACA CATAATATAA GGATAGATAA	2820
ATAAAAACGC ATCTCTGTTT TAAAAACGAA AAAATCGAAA AAGCTTCTCT CTTTCCATA	2880
ATTTTCTACT CAAATTGTGG TACAATTAAG AGTAAGATT TAAGTTAGAA ATGAGACTGA	2940

792

TTGTATGAG AAAATTAAAC AGCCATTCTGA TTCCGATTCTG GCTTAATTTA TTGTTTTCAA	3000
TCGTCAATTT ACTCTTTATG ACCATTATTG GTCGTTTGTT GTATATGCAG GTTTTGAACA	3060
AGGATTTTTA CGAAAAAAG CTAGCTTCAG CTAGTCAGAC CAAGATTACA AGCAGTTCAG	3120
CCCGTGGGGA AATTTATGAT GCTAGTGGAA AACCTTTGGT AGAAAATACG TTAAAGCAGG	3180
TTGTTTCCTT TACGCGTAGC AATAAAATGA CGGCTACAGA CTTAAAAGAA ACAGCTAAAA	3240
AGTTACTGAC TTATGTGAGC ATCAGTTCTC CAAATTTGAC AGAACGCCAG CTGGCGGATT	3300
ACTATTTGGC TGATCCTGAA ATCTATAAAA AAATAGTGGA AGCTCTCCCA AGTGAGAAAC	3360
GCTTGGATTG AGATGGCAAT CGTCTATCCG AATCAGAACT GTATAACAAT GCGGTCGATA	3420
GTGTACAAAC GAGTCAACTA AACTATACAG AGGATGAAAA GAAAGAAATC TATCTTTTTA	3480
GTCAGTTAAA TGCTGTGGA AACTTTGCGA CAGGAACCAT TGCGACAGAT CCTCTAAATG	3540
ATTCTCAGGT GGCTGTATT GCCTCTATTT CAAAGGAGAT GCCTGGCATT AGTATTTCTA	3600
CTTCTTGGGA TAGAAAGGTT TTGGAAACTT CCCTTTCTTC TATAGTTGGG AGTGTATCCA	3660
GTGAAAAAGC TGGTCTCCCA GCGGAAGAAG CAGAAGCCTA TCTTAAAAAA GGCTATTCTC	3720
TAAATGACCG TGTAGGAACC TCCTATTTGG AAAAGCAATA TGAAGAGACC TTACAAGGAA	3780
AACGCTCGGT AAAAGAAATC CATCTGGATA AATATGGCAA TATGGAAAGC GTGGATACAA	3840
TTGAGGAAGG TAGTAAGGGA AACAAATATCA AACTGACCAT TGATTGGGCT TTCCAAGATA	3900
GCGTGGATGC TTTACTGAAA AGTTATTTCA ATTCTGAGCT AGAAAATGGT GGAGCCAAGT	3960
ATTCTGAAGG TGTCTATGCA GTCGCCCTTA ACCCAAAAAC AGGTGCGGTT TTGTCTATGT	4020
CAGGGATTAA ACATGACTTG AAAACGGGAG AGTTGACGCC TGATTCCTTG GGAACGGTAA	4080
CCAATGTCTT TGTTCAGGT TCGGTTGTCA AGGCGGCGAC CATCAGCTCA GGTGGGAAA	4140
ATGGAGTCTT GTCAGGAAAC CAGACCTTGA CAGACCAGTC CATTGTCTTC CAAGGTTTCAG	4200
CTCCCATCAA TTCTTGGTAT ACTCAGGCTT ACGGTTTCATT CCCTATCACA GCGGTCCAAG	4260
CTCTGGAGTA TTCATCAAAT ACCTATATGG TCCAAACAGC CTTAGGTCTT ATGGGGCAAA	4320
CCTATCAACC CAATATGTTT GTCGGCACCA GCAATCTAGA GTCTGCTATG GAGAACTGC	4380
GTTCAACCTT TGGCGAATAT GGCTTGGGTA CTGCGACAGG AATTGACCTA CCAGATGAAT	4440
CTACTGGATT TGTTCCTAAA GAGTATAGCT TTGCTAATTA CATTACTAAT GCCTTTGGGC	4500
AGTTTGATAA CTATACGCCG ATGCAGTTGG CTCAGTATGT AGCAACTATT GCAAATAATG	4560
GTGTTCTGTG GGCTCCTCGT ATTGTTGAAG GCATTTATGG TAATAATGAT AAGGGAGGAC	4620
TGGGTGACTT GATTAGCAA CTGCAACCGA CAGAGATGAA TAAGGTCAAT ATATCCGACT	4680
CCGATATGAG CATCTGCAC CAAGGTTTTT ATCAGGTTGC CCATGGTACT AGTGGATTGA	4740



793

CAACTGGACG TGCCTTTTCA AATGGTGCCT TGGTATCCAT TAGCGGAAAA ACAGGTACAG	4800
CCGAAAGCTA TGTGGCAGAT GGTGAGCAAG CAACCAATAC CAATGCCGTG GCCTATGCCC	4860
CATCTGATAA TCCCCAAATC GCTGTGCGAG TGGTCTTTCC TCATAATACC AATCTAACAA	4920
ATGGTGTAGG ACCTTCCATT GCGCGTGACA TTATCAATCT GTATCAAAAA TACCATCCAA	4980
TGAATTAGAA AGGAAATTAT GCTTTATCCA ACACCTATTG CCAAGTTGAT TGACAGTTAT	5040
TCTAAGTTAC CAGGTATCGG GATTAAGACG GCTACGCGTC TGGCCTTTTA TACGATTGGG	5100
ATGTCTGCTG ATGATGTCAA TGAATTGCA AAAAACTCC TTTCTGCTAA GAGAGAATTG	5160
ACATATTGTT CTATTTGTGG ACGTTTGACA GACGACGATC CTTGTTCTAT CTGTACTGAT	5220
CCGACTCGTG ACCAGACAAC AATTTTAGTT CTTGAGGATA GTAGAGATGT GGCAGCCATG	5280
GAAAAATATCC AAGAATACCA TGGACTCTAT CATGTCCCTC ATGGCCTCAT TTCTCCTATG	5340
AATGGTATCA GTCCGGACGA TATCAATCTC AAGAGCCTTA TGAATCGTCT TATGGATAGT	5400
GAGGTTTCAG AAGTGATTGT GCGGACTAAT GCTACAGCGG ATGGTGAAGC GACTTCCATG	5460
TATCTTTTAC GTTTGCTCAA GCCGGCTGGT ATCAAGGTTA CGGCTCTAGC ACGAGGTCTC	5520
GCTGTGGGAG CGGACATTGA GTATGCGGAC GAAGTGACAC TCTTACGAGC CATTGAAAAT	5580
CGGACAGAGT TGTAAGTGTA GGCAAATTTA CGAACTCCAT TCATTTATAA AAAATCAAAG	5640
AGGCTGAAAA TCGTTCCTAT CGGCCTCTTT TTGTATAGTG TGATGAGTAG GCTCAGGTTT	5700
AAGTTTTAAA AAACCAAGCA AATATGATAT ACTAAAGAGC GAGTATTCTA GTAGAATTAG	5760
GACAAATAAT ATGAAACAAA CGATTATTCT TTTATATGGT GGACGGAGTG CGGAACGCGA	5820
AGTCTCTGTC CTTTCAGCTG AGAGTGTCTG GCGTGCGGTC GATTACGACC GTTTCACAGT	5880
CAAGACTTTC TTTATCAGTC AGTCAGGTGA CTTTATCAAA ACACAGGAAT TTAGTCATGC	5940
TCCGGGGCAA GAAGACCGTC TCATGACCAA TGAAACCATT GATTGGGATA AGAAAGTTGC	6000
ACCAAGTGCT ATCTACGAAG AAGGTGCAGT GGTCTTTCCA GTCCTTCACG GGCCAATGGG	6060
AGAAGATGGC TCTGTTCAAG GATTCTTGA AGTTTTGAAA ATGCCTTACG TTGGTTGCAA	6120
CATTTTGTCA TCAAGTCTTG CCATGGATAA AATCAGGACT AAGCGTGTTC TGAATCTGC	6180
TGGTATTGCC CAAGTTCCTT ATGTGGCTAT CGTTGAAGGC GATGATGTGA CTGCTAAAAT	6240
CGCTGAAGTG GAAGAAAAAT TGGCTTATCC AGTCTTCACT AAGCCGTCAA ACATGGGGTC	6300
TAGTGTGCGT ATTTCTAAGT CTGAAAACCA AGAAGAACTC CGTCAAGCCT TAAAACTTGC	6360
CTTCCGATAT GACAGCCGTG TCTTGTTGA GCAAGGAGTG AATGC	6405

(2) INFORMATION FOR SEQ ID NO: 108:

794

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11309 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

CGAGCTCGGG TACCGGGATT TTAAGGAGTT TGATATGTAT AACCTATTAT TAACCATTTT	60
ATTAGTATTA TCTGTTGTGA TTGTGATTGC AATTTTCATG CAACCAACCA AAAACCAATC	120
CAGCAATGTA TTTGATCCCA GTTCAGGTGA TTTGTTTGAA CGCAGTAAAG CTCGCGGTTT	180
TGAAGCTGTA ATGCAGCGTT TGACAGGGAT TTTAGTCTTT TTCTGGCTAG CCATTGCCTT	240
AGCATTGACG GTATTATCAA GTAGATAAGA AAATAATGGG CAGGACTAGG TCTTTGCCTC	300
TTTTTATTTT TAAAGGATGT TTGAGAAGGT TTTACAGTAA AAGAAAATTA AAAAATCTAG	360
AAAGAAAATA TGAAAGATAG AATAAAAGAA TATTTACAAG ACAAGGGAAA GGTGACTGTT	420
AATGATTTGG CTCAGGCTTT GGGAAAAGAC AGTTCCAAGG ATTTTCGTGA GTTGATTAAA	480
ACCTTGTCCT TAATGGAAAG AAAGCACCAA ATTCGTTTTG AAGAAGATGG TAGTCTGACA	540
TTAGAAATTA AGAAAAACA TGAGATTACC CTCAAGGGGA TTTTTCATGC CCATAAAAAAT	600
GGCTTTGGCT TTGTTAGTCT GGAAGGCGAG GAGGACGACC TTTTGTAGG GAAAAATGAT	660
GTCAACTATG CTATTGATGG TGATACCGTC GAGGTAGTGA TTAAGAAAGT CGCTGACCGC	720
AATAAGGGAA CAGCAGCAGA AGCCAAAATT ATTGATATCC TAGAACACAG TTTGACAACA	780
GTTGTCGGGC AAATCGTTCT GGATCAGGAA AAACCTAAGT ATGCTGGCTA TATTCGTCA	840
AAAAATCAGA AAATCAGTCA ACCGATTTAT GTTAAGAAAC CAGCCCTAAA ATTAGAAGGA	900
ACAGAAGTTC TCAAAGTCTT TATCGATAAA TACCCAAGCA AGAAACATGA TTTCTTTGTC	960
GCGAGTGTTT TCGATGTAGT GGGACACTCA ACGGATGTCG GAATTGATGT TCTTGAGGTC	1020
TTGGAATCAA TGGACATTGT ATCCGAGTTT CCAGAAGCTG TTGTTAAGGA AGCAGAAAGT	1080
GTGCCTGATG CTCCGTCTCA AAAGGATATG GAAGGTCGTC TGGATCTAAG AGATGAAATT	1140
ACCTTTACCA TTGACGGTGC GGATGCCAAG GACTTGGACG ATGCAGTGCA TATCAAGGCT	1200
CTGAAAAATG GCAATCTGGA GTTTGGGGTT CACATCGCAG ATGTTTCTTA TTATGTGACC	1260
GAGGGGTCTG CCCTTGACAA GGAAGCCCTT AACCGTGCGA CTTCTGTTTA CGTGACAGAC	1320
CGAGTGGTGC CAATGCTTCC AGAACGACTA TCAATGGCA TCTGCTCTCT CAATCCCCAA	1380
GTTGACCGCC TGACCCAGTC TGCTATTATG GAGATTGATA AACATGGTCG TGTGGTCAAC	1440
TATACCATTA CACAAACAGT TATCAAGACC AGTTTTCGTA TGACCTATAG CGATGTCAAT	1500

795

GATATCCTAG CTGGCGATGA AGAAAAGAGA AAAGAATATC ATAAAATTGT ATCAAGTATC	1560
GAACTCATGG CCAAGCTTCA TGAACTTTA GAAACATGC GTGTGAAACG TGGAGCTCTC	1620
AATTTTGATA CCAATGAAGC GAAGATTTTA GTGGATAAAC AAGGTAAGCC TGTTGATATC	1680
GTTCTTCGGC AGCGTGGTAT TGCCGAGCGG ATGATTGAGT CTTTATCTT GATGGCTAAT	1740
GAAACAGTTG CCGAACATTT CAGCAAGTTG GATTTGCCCTT TTATCTATCG AATTCACGAG	1800
GAGCCTAAGG CTGAAAAGGT TCAGAAGTTT ATTGATTATG CTTCGAGTTT TGGCTTGCGC	1860
ATTTATGGAA CTGCCAGTGA GATTAGTCAG GAGGCACTTC AAGACATCAT GCGTGTCTGT	1920
GAGGGAGAAC CTTATGCAGA TGTATTGTCC ATGATGCTTC TCGCTCTAT GCAGCAGGCT	1980
CGTTATTTCG AGCACAATCA CGGCCACTAT GGACTAGCTG CTGACTATTA TACTCACTTT	2040
ACCAGTCCAA TTCGTCGTTA TCCAGACCTT CTTGTTTACC GTATGATTCG GGATTACGGC	2100
CGTTCTAAGG AAATAGCAGA GCATTTTGAA CAAGTGATTC CAGAGATTGC GACCCACTCT	2160
TCCAACCGTG AACGTCGTGC CATAGAAGCT GAGCGTGAAG TCGAAGCCAT GAAAAAGGCT	2220
GAGTATATGG AAGAATACGT GGGTGAAGAG TATGATGCAG TTGTATCAAG TATTGTCAAA	2280
TTCGGTCTCT TTGTCCAATT GCCAAACACA GTTGAAGGCT TGATTACAT CACTAATCTG	2340
CCTGAATTTT ATCATTTCAA TGAGCGTGAT TTGACTCTTC GTGGAGAAAA ATCAGGTATC	2400
ACTTTCCGAG TGGGTCAGCA GATCCGTATC CGTGTGAAA GAGCGGATAA AATGACTGGA	2460
GAGATTGATT TTTCATTTCG ACCTAGTGAG TTTGATGTGA TTGAAAAAGG CTTGAAACAG	2520
TCTAGTCGTA GTGGCAGAGG GCGTGATTCA AATCGTCGTT CGGATAAGAA GGAAGACAAG	2580
AGAAAATCAG GACGCTCAAA TGATAAGCGT AAGCATTAC AAAAAGACAA GAAGAAAAAA	2640
GGAAAGAAAC CTTTTTACAA GGAAGTAGCT AAGAAAGGAG CCAAGCATGG CAAAGGGCGA	2700
GGGAAAGGTC GTCCGACAAA ATAAAAAGGC ACGCCACGAC TATACAATCG TAGATACGCT	2760
AGAGGCAGGG ATGGTCCTGA CTGGAACTGA AATCAAGAGT GTACGAGCTG CTCGAATTAA	2820
TCTCAAGGAT GGCTTTGCTC AAGTGAAAAA TGGAGAAGTT TGGCTGAGCA ATGTTTCATAT	2880
CGCGCCTTAC GAAGAGGGCA ATATCTGGAA CCAGGAACCA GAACGTCGTC GTAAACTCCT	2940
GCTCCATAAA AAGCAAATTC AAAAATTGGA ACAAGAGATC AAAGGGACAG GAATGACCTT	3000
AGTTCCCCCTT AAGGTCTATA TAAAAGATGG CTACGCTAAG CTTCTTTTAG GACTTGCCAA	3060
AGGGAAGCAT GACTATGACA AACGGGAGTC TATCAAACGT CGTGAGCAAA ATCGAGATAT	3120
CGCGCGTGTG ATGAAAGCTG TTAATCAGCG ATAAAAAGAG GAATTGAAAA TGGAAAAATT	3180
AGTTGCCTAT AAACGCATGC CTTTGTGGAA TAAACAAACA ATGCCTGAAG CTGTTTCAGCA	3240

796

AAAGCACAAT	ACAAAAGTTG	GGACTTGGGG	GAAAATTACT	GTCTTGAAGG	GAGCTCTCAA	3300
GTTTATTGAA	TTGACAGAAG	AAGGGGAAGT	TCTAGCTGAA	CACCTCTTTG	AAGCAGGGGC	3360
AGACAATCCA	ATGGCCCAAC	CTCAAGCCTG	GCACCGAGTG	GAAGCTGCCA	CAGATGATGT	3420
GGAATGGTAC	TTGGAATTTT	ATTGTAAACC	TGAGGATTAT	TTTGCTAAAA	AATACAATAC	3480
CAATCCTGTT	CATTCAGAGG	TCCTAGAGGC	CATGCAGACA	GTGAAACAAG	GGAAAGCTTT	3540
GGATTTGGGT	TGTGGTCAGG	GGCGTAATTC	TCTTTTCTA	GCCCAGCAAG	ATTTTGATGT	3600
GACGGCTGTA	GATCAAAATG	GACTAGCTCT	TGAAATCTTG	CAAAGCATTG	TGGAGCAGGA	3660
AGATTTGGAC	ATGCCTGTTG	GCCTTTACGA	TATCAATTCA	GCTAGCATTG	AACAAGAATA	3720
TGATTTTATC	GTTTCAACAG	TTGTTCTCAT	GTTCCTACAA	GCCGACCGCA	TTCCAGCTAT	3780
TATTCAAAAT	ATGCAGGAGA	AAACCAGTGT	TGGTGGTTAC	AACCTTATCG	TTTGTGCCAT	3840
GGACACGGAG	GATTATCCTT	GCTCGGTAA	CTTCCCATTC	ACCTTTAAAG	AAGGAGAACT	3900
GGCAGACTAT	TACAAGGATT	GGGAATTGGT	TAAGTACAAT	GAAAATCCAG	GCCATTTGCA	3960
CCGTCGCGAT	GAGAATGGCA	ATCGTATTCA	ACTACGCTTT	GCGACCTTAC	TAGCTAAGAA	4020
AATCAAGTAA	ACACACATGA	AGATTAGGAA	TTTTCTGAT	CTTTTTTCTT	TTTTACGAAT	4080
GATATAGAAA	AGGAGGGAAT	TCATGTTTGT	TGCGAGAGAT	GCTAGGGGAG	AATTGGTAAA	4140
TGTGTTAGAG	GATAAACTTG	AGAAGCAAGC	ATACACCTGC	CCAGCTTGTG	GAGGCCAGCT	4200
CCATTTGCGT	CAAGGACCAA	GTGTACGGAC	GCATTTTGCC	CATAAATCCT	TAAAAGACTG	4260
TGATTTTTTC	TTTGAAAATG	AAAGTCCAGA	ACACCTGGCC	AATAAGGAAT	CCCTCTATCA	4320
CTGTTGAAA	AAAGAGACAA	AGGTTCAATT	AGAGTACCCG	CTTTCAGAAC	TAAACAGAT	4380
TGCGGATGTA	TTTGTAATG	GCAATCTAGC	TCTAGAAGTT	CAGTGTAGTC	CCTTGCCTCA	4440
GAAAGTCCTT	AAAGAGCGAA	GTGAGGGCTA	TCGTAGTCAG	GGTTACCAAG	TACTGTGGTT	4500
GCTGGGTCAA	AAACTGTGGC	TCAAGGAGCG	TTTGACTCGT	CTACAGCAAG	GTTTTCTTTA	4560
TTTCAGTCAA	AACATGGGCT	TTTATGTTTG	GGAATTAGAC	AAGGAAAAAC	AAGTTTTAAG	4620
ACTCAAATAC	CTGATTTACC	AGGATCTCCG	CGGTAAACTC	CATTATCAAA	TCAAGGAATT	4680
TTCTATGGT	CAAGGTAGTT	TATTGGAAAT	ATTGCGTCTT	CCCTATAAGA	GACAAAAAAT	4740
ATCTCATTTT	ACAGTTTCTG	AGGACAAGGA	CATCTGTCGC	TATATCCGGC	AACAACTTTA	4800
TTATCAAAAT	CTCTTTTGA	TGAAAGAACA	AGCAGAAGCC	TATCAAAAGG	GAGAAAATAT	4860
CCTGACTTAT	GGACTGAAAG	AATGGTATCC	ACAAATTCGA	CCAATAGTGG	GCAAAATTTT	4920
CCAGATTGAA	CAAGACTTGA	CTAGCTATTA	TCAGCACTTT	TATACCTATT	ACCAAAAAAA	4980
TCCTCAAAAT	GATTGGCAAA	AGCTTTATCC	ACCAGCCTTT	TATCAGCAAT	ATTTCTTGAA	5040

797

AAATATGGTA GAATAGAAAG GATGGAGGAA TCTAATGGTA TTACAAAGAA ATGAAATAAA	5100
TGAAAAAGAT ACATGGGATC TATCAACGAT CTACCCAAC TACCAGGCTT GGGGAAGAAGC	5160
CTTAAAAGAT TTAACAGAAC AATTGGAGAC AGTAGCCCAG TATGAAGGCC ATCTCTTGGA	5220
TAGTGCGGAT AACCTACTAG AAATCACTGA ATTTTCTCTT GAAATGGAAC GCCAGATAGA	5280
GAAGCTTTAC GCTTATGCTC ATATGAAGAA TGACCAGGAT ACACGTGAAG CTAAGTATCA	5340
AGAGTACTAT GCCAAGGCCA TGACACTCTA CAGCCAGTTA GACCAAGCCT TTTCATTCTA	5400
TGAGCCTGAA TTTATGGAGA TTAGCGAAAA GCAGTATGCT GACTTTTTAG AAGCTCAACC	5460
AAAGCTGCAG GTTTATCAAC ACTATTTTGA CAAGCTTTTG CAAGGCAAGG ATCAGCTTCT	5520
TTCAACAAGT GAAGAAGAAT TATTGGCTGG AGCTGGAGAA ATCTTTGGTT CAGCAAGTGA	5580
AACCTTCGCT ATCTTGACA ATGCGGATAT TGTGTTCCCT TATGTCCTAG ACGATGATGG	5640
TAAAGAAGTT CAGCTATCTC ATGGGACTTA CACACGTTTG ATGGAGTCTA AAAAAGTGA	5700
GGTTCGCCGT GGTGCCTATC AAGCTCTTTA TCGGACTTAC GAACAATTCC AACACACCTA	5760
TGCCAAAACC TTGCAAACCA ATGTTAAGGT GCAAAATTAC CGTGCTAAAG TTCGTAAC TA	5820
CAAGAGTGCT CGTCATGCAG CCCTCGCAGC GAATTTTGTT CCAGAAAGTG TTTATGACAA	5880
TTTGGTAGCA GCAGTTCGCA AGCATTTGCC ACTCTTACAT CGCTATCTTG AGCTTCGTTT	5940
AAAAATCTTG GGGATTTTCA ATCTCAAGAT GTACGATGTC TACACACCGC TTTCATCTGT	6000
TGAATACAGT TTTACCTACC AAGAAGCCTT GAAAAAGCA GAAGATGCTT TGGCAGTCTT	6060
GGGTGAGGAT TACTTGAGCC GTGTTAAACG TGCCTTCAGC GAGCGTTGGA TTGATGTTTA	6120
CGAAAATCAA GGCAAGCGTT CAGGTGCCTA CTCTGGTGGT TCTTATGATA CCAATGCCTT	6180
TATGCTTCTC AACTGGCAAG ACAATCTGGA CAATCTCTTT ACTCTTGTTT ATGAAACAGG	6240
TCACAGTATG CATTCAAGCT ATACTCGTGA AACTCAGCCT TATGTTTACG GGGATTACTC	6300
TATCTTTTTG GCTGAGATTG CCTCAACTAC CAATGAAAAT ATCTTGACGG AGAAATTATT	6360
GGAAGAAGTG GAAGACGACG CAACACGCTT TGCTATTCTC AATAACTTCC TAGATGGTTT	6420
CCGTGGAACA GTTTTCCGCC AACTCAATT TGCTGAGTTT GAACACGCCA TTCACCAAGC	6480
AGATCAAAAT GGGGAGGTCT TGACAAGCGA TTTCCTAAAT AACTCTACG CAGACTTGAA	6540
CCAAGAGTAT TATGGTTTGA GTAAGGAAGA CAATCCTGAA ATCCAATACG AGTGGGCTCG	6600
CATTCCACAC TTCTACTATA ACTACTATGT ATATCAATAT TCAACTGGCT TTGCGGCCGC	6660
CTCAGCCTTG GCTGAAAAAA TTGTCCATGG TAGTCAAGAA GACCGTGACC GCTATATCGA	6720
CTACCTCAAG GCAGGTAAGT CGGACTATCC ACTTAATGTC ATGAGAAAAG CTGGTGTTGA	6780

798

TATGGAGAAG GAAGACTACC TCAACGATGC CTTTGCAGTC TTTGAACGCC GTTTAAATGA	6840
GTTTGAAGCC CTTGTTGAAA AATTAGGATT GGCATAAAAT GGTGAATCG TATAGTAAGA	6900
ATGCTAACCA TAACATGCGT CGTCTGTCG TCAAAGAAGA AATTGTAGAC TTGATGCGTC	6960
AGCGTCAAAA GCAGGTCACA GGTTCCTTGA AAGAATTGGA AGACTTTGCC CGCAAGGAAA	7020
ATATTCCTAT TATCCCCAT GAAACGGTTG CTTATTTCCG TTTCTTATG GAAACCATGC	7080
AGCCTAAAAA TATTCTGGAA ATTGGGACGG CTATCGGTTT TTCAGCTCTC TTGATGGCTG	7140
AACATGCGCC AAATGCTAAG ATTACAATA TTGATCGTAA TCCAGAAATG ATTGGTTTTG	7200
CCAAGGAAAA TTTTGCCAG TTTGACAGTC GCAAGCAAAT CACTCTCCTA GAGGGAGATG	7260
CGGTGGATGT CTTATCTACA CTGACAGAGT CTTATGATTT CGTCTTTATG GATTCTGCCA	7320
AGTCTAAATA CATCGTCTTT CTGCCAGAAA TCCTCAAACA TTTGGAAGTT GGTGGTGTGG	7380
TTGTCTTGA TGATATTTT CAAGGTGGTG ATGTTGCCAA GGATATTATG GAAGTCCGTC	7440
GTGGTCAGCG AACCATTAT CGAGGCCTTC AAAAATTATT TGATGCAACC TTAGACAATC	7500
CAGAACTCAC CGCAACATTA GTGCCTTTAG GAGATGGTAT TCTCATGCTT CGTAAAAATG	7560
TAGCAGATGT TCAACTGTCT GAAAGCGAAT GATTTTCAGA AAAATTTAAG AAAAAATAGT	7620
AAAATAGATA GAGTAACACT TATCTCAAAG GAGTAGACAT GAAGAAAAA TTATTGGCAG	7680
GTGCCATCAC ACTATTATCA GTAGCAACTT TAGCAGCTTG TTCGAAAGGG TCAGAAGGTG	7740
CAGACCTTAT CAGCATGAAA GGGGATGTCA TTACAGAACA TCAATTTTAT GAGCAAGTGA	7800
AAAGCAACCC TTCAGCCCAA CAAGTCTTGT TAAATATGAC CATCCAAAAA GTTTTGTAAA	7860
AACAATATGG CTCAGAGCTT GATGATAAAG AGGTTGATGA TACTATTGCC GAAGAAAAA	7920
AACAATATGG CGAAAACTAC CAACGTGTCT TGTACAAGC AGGTATGACT CTTGAAACAC	7980
GTAAGGCTCA AATTCGTACA AGTAAATTAG TTGAGTTGGC AGTTAAGAAG GTAGCAGAAG	8040
CTGAATTGAC AGATGAAGCC TATAAGAAAG CCTTTGATGA GTACACTCCA GATGTAACGG	8100
CTCAAATCAT CCGTCTTAAT AATGAAGATA AGGCCAAAGA AGTTCTCGAA AAAGCCAAGG	8160
CAGAAGGTGC TGATTTTGCT CAATTAGCCA AAGATAATTC AACTGATGAA AAAACAAAAG	8220
AAAATGGTGG AGAAATTACC TTTGATTCTG CTTCAACAGA AGTACCTGAG CAAGTCAAAA	8280
AAGCCGCTTT CGCTTTAGAT GTGGATGGTG TTTCTGATGT GATTACAGCA ACTGGCACAC	8340
AAGCCTACAG TAGCCAATAT TACATTGTAA AACTCACTAA GAAAACAGAA AAATCATCTA	8400
ATATTGATGA CTACAAAGAA AAATTAAGAA CTGTTATCTT GACTCAAAAA CAAATGATT	8460
CAACATTTGT TCAAAGCATT ATCGGAAAAG AATTGCAAGC AGCCAATATC AAGGTTAAGG	8520
ACCAAGCCTT CCAAAATATC TTTACCCAAT ATATCGGTGG TGGAGATTCA AGCTCAAGCA	8580

799

GTAGTACATC	AAACGAATAG	TCCAAATCAA	TGAGTCAGGG	AAAAAACTCG	ACTTCAGGAA	8640
AAAATGAAGC	AAACATTCCC	ACAATAAAAC	GCATAGTACA	AGGTTTGTAC	TGCCCCCCAA	8700
AAAGTTAGAC	AATTAATTTA	TCCGAAGGAT	TTAGTTCTGT	ATTGCACAGA	GCTAAGTCCT	8760
TTTAGTTTTA	TCTTAATTCT	CTTATTGTTG	TAATAATCAA	TATAGTCTAT	AATGGCTCGT	8820
TCCAATTGAT	TAAGTGATTT	AAATGTTTTC	TCATAGCCAT	AAAACATTTC	GGATTTTAAA	8880
ATGCCAAAGA	AAGATTCCAT	CCTACCGTTG	TCTTGGCTGT	TGCCCTTACG	TGACATGGAT	8940
GCTTGAATTC	CCTTACTCTC	TAGGAAGCGA	TGATAAGAAT	CGTGTTGATA	TTGCCAGCCT	9000
TGGTCACTAT	GGAGAATCGT	ATTCTCGTAG	TGCTTCTCTT	TGAATGCCTG	TTCCAACATT	9060
AACGATCAAT	CAATTTAATC	ATGTACCTAA	GATTAGAATT	GTTTATCCCA	AATTTATTTG	9120
AAAGCTTCTC	TAAGCTATAT	CCTTGTTTTC	TAAGTTCATA	GATCTGAACT	TTATCATCAT	9180
AAGTTAATTT	CATAATAAAA	ACACCCCAAA	AGTTAGATTT	TTTCTGTCTA	ACTTTTGGGG	9240
TGTAGTTCAT	GTACACCTGA	TATGATGCGT	TTTATAATTT	TAAAGACTTT	TTGACCAGCC	9300
TCATTTTTTT	AACTTGATAC	TCAGTGAAAA	GCAAAGATTA	AACTAGGAAG	CTAGCTGTAG	9360
GCTGCTCAAA	GAACAGCTTT	GAGGTGTAG	ATAAACTTG	TGAGGTCACC	AACATATATA	9420
ATGTGAAGCT	GACGTGGTTT	GAATAGATTT	TAGAAGAGTA	TGAGTCTGGA	AGTTTTAATG	9480
GATAATGCAA	GATTCCATAG	AATGGGTAAG	CTAGAGTTCT	TATGTGAAGA	GTTTGGGCAT	9540
AAACTTTTAC	CTTTTCCTCC	CTACTCATCT	TAGTATAGAA	AAGTGAATCT	GAAATAGTAC	9600
ATAACTGCTT	CTAAAACATT	CTTATAAATT	GATTTAAATT	CTCAAATCAT	ATTATTCAGT	9660
TCTTATTTCA	TTTTGTTCTA	CAATCCTGTT	GAGAAGACAC	GTGTTTCATAT	CAAAAAGGTA	9720
TTGGCAAGTT	GCAATACCTT	TTTACGAGGC	TCTGTTGTCT	TATTTTTGTT	TCAACTGACT	9780
ATATCTCCTA	TGGTTCTAGT	TCAGAAGGCT	AGGCTATAAT	TATGATTGAT	AAGAAGTATC	9840
ATTCCAAGTA	TTGGGAGTGA	ATGTTTCAAA	ATCATGGGTT	TCTATAATGG	TCAGGCTGGC	9900
ATTTGCTAGA	CCGCCATCTT	TACGAAGAAG	TGTTTCTTTA	TAGCCTAGGA	GAGTACGAAG	9960
ACTGGCAGTA	AGATTGGCGC	CGTGTCGGAC	AATTAGAATA	CGTTCAGCTG	GACTATCTTT	10020
TAATGATTTG	ATAAATTGGA	TGGTCCGTTG	AGTTGTACTA	TAGAGGGATT	CGGCTCCGAA	10080
CATTCGAGTG	TCAAATTGAG	CAAGATTTGA	ACGAAAAGCC	TGGATTTGTT	GCGGGTAAAT	10140
AGCTTCCAAG	GTTGCAATTT	TCAAACCTTC	TAACTTCCCA	AGTTGCCATT	CACGGAGATT	10200
AGGAACGATT	TCTAAAGAAC	AGGGGGTATA	GAGTTGACTT	TGGATAATCT	CAGCAGATTT	10260
GACCGCTCGA	GGTAAATCAC	TTGAATAAAT	CTGATCAAAA	GGAATTTCTT	TGAGATACTG	10320

800

ACCAAGTCGT TTTAGGGTTT CAATGGATTC AGGAAGAAGA GGAGAATCAC CACTAGCACC	10380
TTGAAAACGA CCTTCTTGGT TCCAGAGGGT ACGACCGTGG CGGACAAAGT AGAGTTTCAT	10440
TACTTGATGT CCTCCAAAAT ATCTACAAAG TCTGCCTTTA CAAAGCTAGC CAAGTCTTGT	10500
GGCGCGACGA TAATGCTGTG TCCGACTTCG CCTGCAGAGA CAATCATTTG ATCCAAATCT	10560
AGAGCAATTT TATCGATAAA AATGGGATAA TTGTGTTTCT GACGAATTCC GACAGGATTA	10620
TTGGCTCCAT GAATGTAACC AGTTGTTTTT TCTAAGTCCT TTTGTGGAAT CATGCTCACT	10680
TTTTTATTGC CAGAAATTTT AGCTAGTTTC TTTTCAGACA AGTCTGAGT GATAGGGACA	10740
ATTCCGATAA TCGGTCCGGT CTGTCTCTCC AAAAGCGCCA AGGTTTTGAA AATCTGATCT	10800
CGTTCATAAC CTTGAGGAAG CTCTCCTTCT AGGGCATTGA TTTGAATCCC CTGATGAGGG	10860
ATAGCTGCTT TAGATAGGAT TTGTTCCACC AATGTTTTTT TGATTTTAAC TTTTTTTGCC	10920
ATTATTTATA TTTATCCTCC AATTGACTCA TCCAAATACC AAGCCAGATT CCCAGCGCAA	10980
AGAAGAAGGC GATGATGACA TAACCGACAA GTGAAAGTCC TGTGTATTGG ATACTTTCAG	11040
CGTTTCCTGC ATTTGGAATT AAGATCAAAA GGGTACTTGA TAGGACGATA CCGATGATGA	11100
AATGATAGAC GAACTGTTTA CGGAGTTCTT CTAGTTCTCC GTCCGTCCAA GCGTAGGCCA	11160
CTTCTTCTTT CTTCCTTTA CCTTTGGACA TCTTGTAAG AGGTGGGAGG GCAATATAGA	11220
CATGACCTGC CTCGACTAGC GGACGCATGT AACGGTAGAA AAATGTCAAG AGCAAGGTCT	11280
GGATATGGGC ACCGTCGGTA TCCGCATCG	11309

## (2) INFORMATION FOR SEQ ID NO: 109:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5548 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

CCATAGTCTA ACAAGTCTTT GTAAAGGTTT ATCCCTGATT CATGTAAAGA TTGTGTAAAG	60
AATCAAAAAA AGCCACTTTT GAAAAATGGC TGCTCCTAAA AATAGCTTTA AAAATTATTA	120
GTCTGTGCG AAAGATTGGT TAGGAAGAAA AATCGTGAAG CAACTGCCTC TGCCAAGCTG	180
ACTCGTCACC GTGACTTGGC CACCTAATAA TTGACTGAGT TCTTTGACAA TGGCAAGGCC	240
AAGACCAGTG CCACCAGTTT GTCTGCTTCG ACCTTTATTA ACTCGGTAAA AACGTTCAAA	300
AATACGATCC TGCTCTAATT GACTAATACC AATCCCTGTA TCTGATACAG AAATCTTAAT	360
GCCTTCGTTT ACCTTTTGGG TCTTGACCTC AATTTTCCC CCTTGTTCAG TGTAACGGAT	420



801

GGCATTGGAT AAAAGATTGA GTAAGATTTG GGAAAGTAAT TGACTATCTG ATACGAGGGT	480
GACATCATCT GGCACCTGCA CCTTTAGCTG TAAATCCTTC TTCTTGAGCT GAGGTTGCAA	540
GCTTTGAGTC AAATCCTGTA CAAATTCTGC CAAAGAAAGG GTCGTCCATT GTATAGGCAT	600
TTGTTGAGCC TTAGATAAGG TAAGAAGATG CTCAACAATA TGCTCAAGAC GCAAACCTTC	660
TTTGTAATA ATGTCTAGAA AGTCATCCTT GAGCGCTTCT TCTTCAGCTG ACATCCCCTT	720
AATGGTTTCA GCAAAGCCCT TAATCGAAGT AACTGGTGTC CTCAATTCAT GGGAGGCATT	780
TGAGACAAAG GCTAAATTTA ACTTTTCATA AGTTCTAATC GTTGTTAAAT CATATAGCAA	840
GACGAGCACA GCTTCCACAG ATTGGGTGGG GCTAAAAACG GGAAGTCTG TCACCTCTAA	900
AATCAAGTCA CCCTCATGAA ACCCACTTAC TTCTTGTTTT AACCTTGTTT TTTGATCAAA	960
GGCTTGGTGA ACTAAATTCC GAATATCCAT CCGTTTGAGG TCATCAAGTG AACTTATGTC	1020
GCCGTCCACA TCGGGAAAAT AATGAGGCAG AGAGCGACTG GATAATAACA TCTGACCTTG	1080
AGCGGAAACT AAAAACGTCC CCATGGTTAG GTGCGACAGA AGAACCTCCA TTGTTTCGGC	1140
TAGATCCTTG TATTGCTGAT CCTGTTGGGA GACTTTGGTT TTTAGGCCAG ACACATACTG	1200
AGCCAAAGAC TTTAAGTCTT CTTGCCCTTT TTCTAAAAAG TATTCACTAC TGGTCAAGAG	1260
AGGTTGGTGC AAGGTCTCAA AAGCAACTTC CCATTTCCAA AGGCAAAAGA GCCAGTAGCC	1320
ACCTAGTCCC AAAGAAAGGG CTAGAAGAAA GAGACCGATG CCTTTACTGA TCCAAGTTAA	1380
TGCCATCCCT GCAATCAGAA TGAGGCTAAC ACTTAGATTG ACTAGCCAAA ATTGAAGGTA	1440
GCGTTTCATC TATAACTCCT TGAAGTTATA ACCATAACCC CGAATGGTTC GAATAAATTG	1500
AGGGGCTTTA GGATTGTCTT CAATTTTTC CTTCACTTA CCAATATGAA CGTCCACCAA	1560
ACGTGTTTCC TGCCCAAAGT CATACCCCA GATACGTTCC AAAAGACGCT CTCTAGTCAG	1620
TGTCATGTTG GGATGTTTCA TAAGATAGAG CAAGAGTTCA AATTCTTTTG GGGTCAAAC	1680
CAGTAACTTA TTCGCCTTGT AGACTTCATG ACGCTCAGGG TATACTTTCA AGGTCCCAAA	1740
TAGCCAAGAA TCGTCAGCGA TATTATCTGA ATCATCTCCT TCTTGTTCTC CTTTAGTTTCG	1800
CCTGAGGACA GCCTTGACAC GCGCCAGCAA TTCTCTAGGG CTAAAAGGCT TGGTCAGGTA	1860
GTCATCAGCC CCTAATTCCA AGGCCAAAAC CTTATCAAAT TCATCACTTT TCGCAGAAAC	1920
CATCATAATT GGAGTTTGA CGCCTTTGGC TCTCAGCCGC TTACAACTT CCATGCCATC	1980
TAATTGTGGT AACATGATAT CAAGCAAGAT AAAATCAAAG GCTTCTGTTT CTGCCAAAGC	2040
TAAGGCCTTC CGTCCATTG TCACCAATTG AGTAGAAAAG CCTTCCTTAC TTAAATGGTA	2100
GTCAAGCAAT TTCAGAATGT GTTCTTCATC ATCCACTAAT AAGACTTGTT TTGTCATCTA	2160

802

TTATCTCCTA TTGGTAACAT TATAACACAA TTATCAGAAA TCCTAACATT GCTAAATCAG	2220
ATTAAATTTG CCTATCAAGA CTAGTATCTG GTCAAACGCT CAATCATCTC CTTGTGCTCT	2280
GGATAGGTCG CCAGTAGATC TACCCTTTCA AATAATTCAA AATCCTCAAA TTCAAAACCA	2340
GGAGCAACAA GACAAGAAAC CAGAGCATCA TCCTTATCAA CTGTTGATCC CCAAATAGTG	2400
CCCTTAGGAA CACAGTAGTG AAGTTGTTGC CCTTTGGATA TGTCCAGGCC TAAAGTGACT	2460
GCTTCGTAGT GACCATCTGC TGTAATCATG TGAACAGTAA GTGGGGATCC TGCATGAAAA	2520
TACCAGATTT CATCTGCTGT CAATCGGTGA AAATGTGAAG GATTTCGTTT TTCTAATAAG	2580
AAATAAATAC TGGTATAAAG CGCCCTTCCC TTACCAGCAA GGTTTATAGT GTCTGAAGCT	2640
TTTTTTGTTT GTCTAAAATA GCCACCTTCA ATATGGGGAG CTAACCTAG AGTTCCTATC	2700
AAGTCTTCTT TATCCGTCGG AGCCAATGGG TTGAAGTAAC TCTTGTTCAA AGTGGTTTTA	2760
CGATTTCAGG AACTCCTCTC AGTTCGTAGG ACACGGTAAT GATTGATGCG ACGGAAGTAC	2820
AAATCAATCG CCCTAAAAAA AGAATTAGCG AATGATTCTG GTAAAAAAA TGCCACGCTA	2880
TGAAGGCTCA AGCGATTGTC ACAAGTCAAG GGAGAATTGT TTCTTTGGAT ATCGCTGTGA	2940
ACTATTGTCA TGATATGAAG TTGTTCAAAA TGAGTCGCAG AAATATCGGA CAAGCTGGTA	3000
AAATCTTGCC TGACAGTGGT TATCAAGGGC TCATGAAGAT ATATCCTCAA GCACAAACTC	3060
CACGTAAATC CAGCAAACTC AAGCCACTAA CAGTTGAAGA TAAAGCCTAT AACCATGCGC	3120
TATCCAAGGA GAGAAGCAAG GTTGAGAACA TCTTTGCCAA AGTAAAAACG TTAAAAATGA	3180
TTTCAACAAC CTATCGAAAT CATCGTAAAC ACTTCGGATT ACGAATGAAT TTGATTGCTG	3240
GCATTATCAA TCATGAACTA GGATTCTAGT TTTGCAAGAA GTCTATTATT TGGTTAGGTG	3300
AATTAGTGAA GCGTTTAGGC AAGTGTCTCT GGTACGACG TCATGGACTC TAAATCGATT	3360
ATATTTAGGG GTCATGACTA GTGAAGCAGT TAGCTAGTTC GCATATAAGC GGCTAGCGTC	3420
TAACAATTAG GAACTTTAGT TCCAATAACT TTAAGATTAC GACGTTTTAG GACATAAATC	3480
GATCATATTT ATGTCCTAAA ACTAGTGAAG CGCCTAGCCA AAGTCCGAAT AGGATTTCGC	3540
GTTAGTTACT TAGATTGCTT TGCAATCAAG TAACTTTGGC GATTTACATC TTCTCTGGCG	3600
CTTCTACTCC AAGCAAGCGA AGGGCTTCTT TGAGAACGAC TGCGGTTGCG TAGCTGAGGG	3660
CTAGACGGCT GTCGCGTTCT GGGCTTTCAT CCAAGATACG TGTATGTGCA TAGTATTGT	3720
TAAAGGATTG AGCCAGGCTA ATTGCAAATT TAGCAATGAT AGAAGGTTCA AAGTTATCTG	3780
CCGCACGGTT GATAATACGT GGGGAAGTCTT GAATGAGTTT AATGATTTCC CAGCTTTCAG	3840
TATCATTCOA GCTATAGTTG CCAGCTGTTT CTGGTTTGAA ATCGGCTTTG CGTAAGATAG	3900
ATTGGATACG AGCGTAGGCA TATTGAACGT AAGGTCCAGT TTCACCCTCG AAGGATACCA	3960

803

TAGCCTCTAG GTCGAAGTCG TATCCATTTG TACGGTCGGT TTTGAGGTCA TAGAATTTAA	4020
TGGCTCCAAT CCCAACAGCA TGTGCTACTT GGTCTTTGTT TTCTAGTTCA GGATTTTATAG	4080
CCTCGATTTG GACCTTGCCA CGGCTAACAG CCTCTGCAAC AGTAGGCTCT AGCAAGATGA	4140
CATTCCCTTT ACGAGTAGAG AGTTTCTTCC CTTCTTTTGT AACCAAACCA AAAGGAACGT	4200
GAGTAATGTC GTCACTCCAG TCGTAGCCCA TCTCTTGCAA GACAGCTTTG AGCTGTTTAA	4260
AGTGGGCAGA TTGTTCTTGA CCAACGACAT AGATAGATTT AGCAAATTGG TATTCGTTTT	4320
TACGGTAGAG GGCTGCAGCC AAGTCACGTG TGATATAGAG AGTTGCACCA TCAGACTTCT	4380
TGATGAGGGC TGGATGTTCA ATTCCATATT TCTCAAGATT CACAACTTGG GCACCTTCTG	4440
ATTCAAGAAG TAGTCCTTTT TCAGAAAGAA TGTCTACAAC TGCATCCATC TTATCATTGT	4500
AGAAGGCTTC TCCGTTATAG CTGTCAAATT CAACCTTCAA TTCATTGTAA AGGCGGTAA	4560
ATTCCACTAA ACTTTCATCG CGGAACCATT GCCAAAGAGC GAGAGCTTCC TCATCTCCAT	4620
TTTCAAGTTT ACGGAACCAT TCGCGCGCTT CTTTCATCCAA GCTAGGGTCA TTTTCAGCTT	4680
CAGCGTTGAT GCGGACATAG AGTTTAAGGA GTTCATCGAT TGGATGAGCT TTTACAGCTT	4740
CTTCGTCGCC CCATTTTTTG TAGGCAACAA TCAACATCCC AAATTGTTTA CCCCAGTCTC	4800
CCAAATGGTT GACCTTGACC GTTTGATAAC CGATTTTTTG GAAAATATGT GACAAGCTAT	4860
CTCCGATAAC AGTTGAACGC AGGTGGCCAA TAGAAAATGG TTTAGCGATA TTCGGACTAG	4920
ACATGTCGAT AACAACATTT TCTTGTTTAC CAATATTTTG GTCAGCATAG TGTCTTTTTT	4980
CAGTGGTAAC AGCTTGCAAT ACTTGAGCAG AAATGGCAGA TTTATCAAGG AAAAAGTTAA	5040
CGTAAGGTCC TGTGCGACA ACTTTTCAA AGGCTTGGCT GTTCATTTTT TCAGCCAGTT	5100
CAGCCGCAAT CATTTGTGGT GCTTTACGTT CGACTTTTGC AAGAGAAAAA GCAGGGAAAG	5160
CAATGTCTCC CATTTCTGAG TTTTtagggg TTTCCAGTAA CTTTAAAATA GCCTCTTGGT	5220
CCAGGCTATC AATGATGCTA GATAATTCGC TAGCAATCAA TTCTTTTGTA TTCATTAAGA	5280
GCTCCTTTTT GGACTTTTCT ACTATTTTAT CACAATTTTA AAGAAAGAAG AAAAAATTTT	5340
TGAAATCTCC TGTTTTTTTG GTATAATATG GTTATAAATA TAGTTATAAA TATGCACGCA	5400
AGAGGATTTT ATGAGAAAAA GAGATCGTCA TCAGTTAATA AAAAAATGA TTACTGAGGA	5460
GAAATTAAGT ACACAAAAAG AAATTCAAGA TCGGTTGGAG GCGCACAATG TTTGTGTGAC	5520
GCAGACAACC TTGTCTCGTG ATTTGCGG	5548

(2) INFORMATION FOR SEQ ID NO: 110:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3132 base pairs

804

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

TACCCGGTAG TCTTAGCAGA CACATCTAGC TCTGAAGATG CTTTAAACAT CTCTGATAAA	60
GAAAAAGTAG CAGAAAATAA AGAGAAACAT GAAAATATCC ATAGTGCTAT GGAAACTTCA	120
CAGGATTTTA AAGAGAAGAA AACAGCAGTC ATTAAGGAAA AAGAAGTTGT TAGTAAAAAT	180
CCTGTGATAG ACAATAACAC TAGCAATGAA GAAGCAAAAA TCAAAGAAGA AAATTCCAAT	240
AAATCCCAAG GAGATTATAC GGACTCATTG GTGAATAAAA ACACAGAAAA TCCCAAAAAA	300
GAAGATAAAG TTGTCTATAT TGCTGAATTT AAAGATAAAG AATCTGGAGA AAAAGCAATC	360
AAGGAACTAT CCAGTCTTAA GAATACAAAA GTTTTATATA CTTATGATAG AATTTTAAAC	420
GGTAGTGCCA TAGAAACAAC TCCAGATAAC TTGGACAAAA TTAAACAAAT AGAAGGTATT	480
TCATCGGTTG AAAGGGCACA AAAAGTCCAA CCCATGATGA ATCATGCCAG AAAGGAAATT	540
GGAGTTGAGG AAGCTATTGA TTACCTAAAG TCTATCAATG CTCGGTTTGG GAAAAATTTT	600
GATGGTAGAG GTATGGTCAT TTCAAATATC GATACTGGAA CAGATTATAG ACATAAGGCT	660
ATGAGAATCG ATGATGATGC CAAAGCCTCA ATGAGATTTA AAAAAGAAGA CTTAAAGGC	720
ACTGATAAAA ATTATTGGTT GAGTGATAAA ATCCCTCATG CGTTCAATTA TTATAATGGT	780
GGCAAAATCA CTGTAGAAAA ATATGATGAT GGAAGGGATT ATTTTGACCC ACATGGGATG	840
CATATTGCAG GGATTCTTGC TGGAAATGAT ACTGAACAAG ACATCAAAAA CTTTAACGGC	900
ATAGATGGAA TTGCACCTAA TGCACAAAT TTCTCTTACA AAATGTATTC TGACGCAGGA	960
TCTGGGTTTG CGGGTGATGA AACAATGTTT CATGCTATTG AAGATTCTAT CAAACACAAC	1020
GTGATGTTG TTTCGGTATC ATCTGGTTTT ACAGGAACAG GTCTTGATAG TGAGAAATAT	1080
TGGCAAGCTA TTCGGGCATT AAGAAAAGCA GGCATTCCAA TGGTTGTCGC TACGGGTAAC	1140
TATGCGACTT CTGCTTCAAG TTCTTCATGG GATTTAGTAG CAAATAATCA TCTGAAAATG	1200
ACCGACACTG GAAATGTAAC ACGAACTGCA GCACATGAAG ATGCGATAGC GGTCGCTTCT	1260
GCTAAAAATC AAACAGTTGA GTTTGATAAA GTTAACATAG GTGGAGAAAG TTTTAAATAC	1320
AGAAATATAG GGGCCTTTTT CGATAAGAGT AAAATCACAA CAAATGAAGA TGAACAAAA	1380
GCTCCTAGTA AATTAAAAAT TGTATATATA GGCAAGGGGC AAGACCAAGA TTTGATAGGT	1440
TTGGATCTTA GGGGCAAAAT TGCAGTAATG GATAGAATTT ATACAAAGGA TTTAAAAAAT	1500
GCTTTTAAAA AAGCTATGGA TAAGGGTGCA CGCGCCATTA TGTTGTAAA TACTGTAAAT	1560

805

TACTACAATA GAGATAATTG GACAGAGCTT CCAGCTATGG GATATGAAGC GGATGAAGGT	1620
ACTAAAAGTC AAGTGTTTTC AATTTTCAGGA GATGATGCTG TAAAGCTATG GAACATGATT	1680
AATCCTGATA AAAAACTGA AGTCAAAAGA AATAATAAAG AAGATTTTAA AGATAAATTG	1740
GAGCAATACT ATCCAATTGA TATGGAAAGT TTTAATTCCA ACAAACCGAA TGTAGGTGAC	1800
GAAAAAGAGA TTGACTTTAA GTTTCACCT GACACAGACA AAGAACTCTA TAAAGAAGAT	1860
ATCATCCTTC CAGCAGGATC TACATCTTGG GGGCCAAGAA TAGATTTACT TTTAAAACCC	1920
GATGTTTCAG CACCTGGTAA AAATATTAAA TCCACGCTTA ATGTTATTAA TGGCAAATCA	1980
ACTTATGGCT ATATGTCAGG AACTAGTATG GCGACTCCA TCGTGGCAGC TTCTACTGTT	2040
TTGATTAGAC CGAAATTAAA GGAATGCTT GAAAGACCTG TATTGAAAAA TCTTAAGGGA	2100
GATGACAAAA TAGATCTTAC AAGTCTTACA AAAATTGCCC TACAAAATAC TCGCGGACCT	2160
ATGATGGATG CAACTTCTTG GAAAGAAAA AGTCAATACT TTGCATCACC TAGACAACAG	2220
GGAGCAGGCC TAATTAATGT GGCCAATGCT TTGAGAAATG AAGTTGTAGC AACTTTCAAA	2280
AACACTGATT CTAAAGGTTT GGTAAACTCA TATGGTTCCA TTTCTCTTAA AGAAATAAAA	2340
GGTGATAAAA AATACTTTAC AATCAAGCTT CACAATACAT CAAACAGACC TTTGACTTTT	2400
AAAGTTTCAG CATCAGCGAT AACTACAGAT TCTCTAACTG ACAGATTAAA ACTTGATGAA	2460
ACATATAAAG ATGAAAAATC TCCAGATGGT AAGCAAATTG TTCCAGAAAT TCACCCAGAA	2520
AAAGTCAAAG GAGCAAATAT CACATTTGAG CATGATACTT TCACTATAGG CGCAAATTCT	2580
AGCTTTGATT TGAATGCGGT TATAAATGTT GGAGAGGCCA AAAACAAAAA TAAATTTGTA	2640
GAATCATTTA TTCATTTTGA GTCAGTGGAA GCGATGGAAG CTCTAACTC CAGCGGGAAG	2700
AAAATAAACT TCCAACCTTC TTTGTCGATG CCTCTAATGG GATTTGCTGG GAATTGGAAC	2760
CACGAACCAA TCCTTGATAA ATGGGCTTGG GAAGAAGGT CAAGATCAAA AACACTGGGA	2820
GGTTATGATG ATGATGGTAA ACCGAAAATT CCAGGAACCT TAAATAAGGG AATTGGTGGA	2880
GAACATGGTA TAGATAAATT TAATCCAGCA GGAGTTATAC AAAATAGAAA AGATAAAAAAT	2940
ACAACATCCC TGGATCAAAA TCCAGAATTA TTTGCTTTCA ATAACGAAGG GATCAACGCT	3000
CCATCATCAA GTGGTTCTAA GATTGCTAAC ATTTATCCTT TAGATTCAAA TGGAAATCCT	3060
CAAGATGCTC AACTTGAAAG AGGATTAACA CCTTCTCCAC TTGTATTAAG AAGTGCAGAA	3120
GAAGGATTGA TT	3132

(2) INFORMATION FOR SEQ ID NO: 111:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14672 base pairs

806

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

CGAGATTCTTCT TAAATGAAC TACGTGAAAT CTACCCATCA TCCAGATCTG GATATTCTCT	60
CCTATCTATA AGTAAAGTTT TAGGAGATTT TAATATAAGT TCTCATGCTT TTAAAGCTTC	120
GGTAAGAGAT TAAAACCGC TCAGTTTCCC ACTCATTGTC TTCTGGGAGA GTTCTCATTT	180
TATTATTCTT GAAAAAATTA GTAAAAACAA GTTTTATATT TTAGATCCTG CAAAAGGCAG	240
GCAGAGAATG TCAATAAGTG AATTTGAAAG GCATTATTCA AATATCATTT TAACATTTAA	300
AAAGTTAGAT AGCTTTATGT CTCGTAAAGA TAATAAGAAG TCGCCTGTTT TAAAGTATTT	360
TTTAAAGTAT AGGAATAAGC TAGGGATTTT ATTTTTGTA ACAGCATTAT TGATGTAAAT	420
ACAATCATTG GTACCTATAG CTAATAGATA CATAATTGAC ACGAATTTCA AGGACGATTC	480
GTATTCGTCT AGAATGTTAT TTAATATATT ATTTATATTT ACTGTTTCAT TCTCACTAAT	540
GTATTTATTA AGACAGATAT ATGTTGCATC CTTAAATAT ATAATGGATA AAGAGATTAG	600
CTATGATTTT ATGAAACATT TGATATATTT ACCTTACAGT TTTTATGAAA AACGTACTTT	660
AGGGGATATA CTTTTAGAG CTAACCTCTAT TGTTTATATA AGAGAAATAC TATCAAATAA	720
TTTTATAGCA GCTATACTTG ATTTGTTAAT GATTGTGGTT TATGCTGTGG TTTTATTTAG	780
CTTTTCTAAG TACATGGTAA TCTTTTAAAT ATCACTAAGT CTAGCTCTAT CTATTGTAAT	840
GTATCCAATC ATAAAAATCT CAAAAAATTT AATTGATAAA AATATAAAAG AAAAGGTAA	900
TGTTCAAAT ATTACTTCCG AAGTAATTC TAAAAATAGT GATATTAAGC TAACTGGAGA	960
AGAGGAATTT TGGATTAAAC AATGGGATAA TTTAATACA AAACAGCTCA TCATAGGTCTG	1020
AAAACCTGAT ATACATTTAT CAATTGTTAG TAGTATAACG AATGTTTTAC AAATTATTCT	1080
CCCTGTTTTG ACCCTTATTG TAGGTGTAAA TATAAAAACA TTCGAACAAT TCACGTTAAG	1140
ACAAATTGTA GCAATAAGTA CAGTCTCACC ATACTTTATT TCTCCTATAA TTTCTTTAAG	1200
TGATAACTAT ATACAAATTA TGTATTAAA GGGATATTTT TTAAGAATAG AGGATGTGTT	1260
TAATACTAAA TCCGAATTA TCCAGAAAAG AGTCAGTCAA GATATAAAAT TTGATAAAAA	1320
AATAGAATTA AAAGATATTT GGTATAAATA TGGATTATTT GATGATTATG TTTTGAAAGG	1380
AATAAATGTT ACTATTAAAA AAGGAGAAAC TGTGCTATT GTTGGAGAAT CAGGTTGAGG	1440
TAAGAGTACA TTAGCTAAAA TTTTATTAGG TTTATTAGAA CCTAATATTG GTTCAATAGA	1500
AGTTGATGGA GTAGAAAAAG AAGAAATTGG TCAACATTG TATAGAAAAG TTTTGGAGC	1560

807

AGTGTTACAA AATCAACCC TAAGTTATGG TACCTTAAGA GAGAATTGA CATTTGGACA	1620
CTTTGTTTCA GATGAAGAAT TAATGACAAA TCTAAATTCA ATTGGTCTTA GCAATGTAGT	1680
TAAATCTTTA CCTCTTGGAT TAGAGACAAT CATCGCTGAA GAAGGTAATA ACTTTTCTGG	1740
AGGGCAGCAG CAAATGATAC TTTTAGCTCG TTGTCTTTTG TCGAAACCTT CGGTAGTTGT	1800
TTTGGACGAA GCAACAAGTA GTTTAGATAA TTTATCTCAA CAAATTACAA CTTCTTACTT	1860
AAGTGAAATC GGTACCACTA AGATTTTAAT TGCCCATCGA CTAGATACTA TCAAGTCTGC	1920
AGATAAGATC TTAGTAATGC ATAATGGTGA AATTGTAGAG ATTGGGACCC ATAGAGAACT	1980
TCTTGAAC TA GGAGGCATTT ATAAGCAATT GTATTCAAAT AATTAGTTTT TGATTAAAAG	2040
GGTAAATTTA TGAAGATTAT GAAAAAAAAA TATTGGACTT TAGCGATATT ATTCTTTTGT	2100
TTGTTCAATA ATTCTGTTAC TGCTCAAGAA ATACCTAAAA ATCTTGATGG CAATATAACT	2160
CACACTCAGA CTAGCGAAAG TTTTCTGAA TCTGATGAAA AACAGGTTGA CTATTCTAAT	2220
AAAAATCAAG AAGAAGTAGA CCAAAATAAA TTTCGTATTC AAATCGATAA GACAGAATTA	2280
TTTGTAACAA CAGATAAACA TTTAGAAAAA AACTGTTGTA AATTGGAACT TGAACCACAA	2340
ATAAATAACG ATATTGTTAA CTCTGAAAGT AATAATTTAC TAGGCGAAGA TAATTTAGAT	2400
AATAAAATTA AGGAAAATGT TTCTCATCTA GATAATAGAG GAGGAAATAT AGAGCATGAC	2460
AAAGATAACT TAGAATCGTC GATTGTAAGA AAATATGAAT GGGATATAGA TAAAGTTACT	2520
GGTGGAGGCG AAAGTTATAA ATTATATTCT AAAAGTAATT CTAAAGTTTC AATTGCTATT	2580
TTAGATTTCAG GAGTCGATTT ACAAATACT GGATTACTGA AAAATCTTTC AAATCACTCA	2640
AAAACTATG TCCCCAATAA AGGATATTTA GGAAGAGAGG AGGGAGAGGA AGGAATAATA	2700
TCAGATATTC AAGATAGATT AGGTCATGGT ACGGCTGTTG TAGCTCAAAT TGTAGGGGAT	2760
GACAAATTA ATGGAGTAAA TCCTCACGTT AATATTAACG TCTATAGAAT ATTTGGTAAG	2820
TCGTCAGCTA GTCCAGATTG GATTGTAATA GCAATTTTTG ATGCTGTAGA TGATGGCAAT	2880
GATATTATCA ATCTTAGTAC TGGACAATAT TTAATGATTG ATGGAGAATA TGAGGACGGA	2940
ACAAATGATT TTGAAACATT TTTGAAGTAT AAAAAGGCTA TTGATTACGC GAATCAAAAA	3000
GGAGTAATTA TAGTAGCTGC ATTAGGGAAT GACTCCCTAA ATGTATCAAA TCAGTCAGAT	3060
TTATTGAAAC TTATTAGTTC ACGCAAAAAA GTAAGAAAAC CAGGATTAGT AGTTGATGTT	3120
CCAAGTTATT TCTCATCTAC AATTTCCGTC GGAGGCATAG ATCGCTTAGG TAATTTATCA	3180
GATTTTAGCA ATAAAGGGGA TTCTGATGCA ATATATGCCG CTGCAGGCTC AACATTATCT	3240
CTTTCAGAAT TAGGACTTAA TAACTTTATT AATGCAGAAA AATATAAAGA AGATTGGATT	3300

808

TTTTCGGCAA CACTAGGAGG ATATACGTAT CTTTATGGAA ACTCATTGTC TGCTCCTAAA	3360
GTTTCTGGTG CGATTGCAAT GATTATTGAT AAATACAAAT TAAAAGATCA GCCCTATAAT	3420
TATATGTTTG TAAAAAATT CTGGAAGAAA CATTACCAGT AAAAAATGGT ATAAAAAGTG	3480
TAAATATACC AAACGTATTG AGATATGATT TGAATATGTT ACAATTAGAA TATAAAAAATG	3540
AACAAAGTTG GGATAGTTTC ATAGATAATG TTAATTTAAT TGAGTTGGAA GAGAGAATTC	3600
AAACTACTAT TGGAATTAAA CAAATAAACA CACACAATAT TATTACTATT GCCCGAGAAG	3660
GGTACTCTCA AAATTATTTA CCTAACACTT CAGAAAATAC ATATAATTCA TTACAAGTCA	3720
GTTTAGTTGG AGTATTACTA CTTTTTATAA GTATGGTAAA TATTTTATGG GCTAAAAAAA	3780
GTAAATGAAA ATAAAAATTG GAGCCCTCTG AAAAAAGTAAG TCCTACAGTT CAACTAAAAT	3840
GAGTCAAAAG ATGAATCACC TTGATGTAGG GGAGTTTGTC TTATTGCTGC CTGAACACCT	3900
CCGTTTCAGAG GAAGAACATT ATAAATCTGT TTTTGAAGAC GACTTAACCA GTCCGATATC	3960
TAGTCAAGAT GAACGACAGC AAATGACTGC TACGGTAGGT TATTTAGAAT CAGGTCAGGA	4020
TCGTTTTGTG TATAATACGA CCCCTATTTT TTACCAGCAG TTTTGAAGAT ATCCAATCAT	4080
CATTGTTATA ACACCCCAAT CAACTGGTCC ACAGTCCATT TTGTTTTGGA TAGACGCAGT	4140
ACAGAACTAC GTTCTCTTTA ATCAATTGTC TGATGCCCAG GAGCTTATCC AGAGACAAGG	4200
CATTGAAAAT TGGGTCTCAG AAATGCAAAC AGGTTACCAC AACTACATCA CATTATTGGA	4260
TAATATCCAG AGGGAACGTT GGGTAATGCT AGCAGGAGCT GTGCTTGGGA TTGCAACTTC	4320
AATCTTGTTG TTAAACACTA TGAATAGGCT CTACTTTGAA GAATTTAGAC GTGCCATTTT	4380
TATCAAACGC ATTGCAGGTC TCAGGTTCTT AGAAATCCAT CGCACTTATC TCTTTGCTCA	4440
ACTGGGTGTG TTTTACTGG GATTTGTTGC GAGTGTATTT CTTCAGGTAG AGATAGGAGT	4500
TGCTTTCTTA GTCTTGTTAC TCTTTACTGG TCTATCTCTT TTACAGTTAC ATGTCCAAAT	4560
GCAGAAAGAA AACAAGATGT CCATGCTTGT TTTGAAGGGA GGTTAATATG ATTGAACTTA	4620
AACAGGTGAG TAAATCTTTT GGAGAACGAG AGTTATTTTC GAATCTTTCA ATGACATTTG	4680
AGGCTGGAAA AGTCTATGCC TTAATTGGTT CAAGTGGTAG CGGAAAAACA ACCTTGATGA	4740
ACATGATTGG GAAATTAGAA CCTTATGATG GGACGATTTT TTACCGAGGT AAAGACTTGG	4800
CCAATTATAA ATCAAGTGAT TTTTCCGTC ACGAATTGGG CTACCTCTTC CAGAACTTTG	4860
GCTTAATTGA AAACCAAAGT ATTGAAGAAA ACCTTAAGCT AGGTCTCATT GGTCAAAAGT	4920
TGAGTCGGTC GGAACAGCGG TTGAGGCAGA AGCAGGCTTT AGAACAGGTC GGCCTGGTTT	4980
ATCTTGACCT AGATAAGCGC ATCTTTGAGT TATCGGGCGG AGAATCGCAA CGGGTTGCCT	5040
TGGCAAAAAT TATCTTAAAG AATCCACCCT TTATTCTGGC AGATGAGCCA ACAGCTTCAA	5100



809

TAGACCCAGC AACCTCTCAG TTGATTATGG AGATTTTGCT ATCTCTTCGA GATGATAATA	5160
GGCTAATCAT TATCGCAACA CATAATCCGG CAATTTGGGA GATGGCTGAT GAACTGTTCA	5220
CGATGGATCA TCTGAAATAA AAATCCTTGT TTTAATTGTC ACGATGAGTT ACTGAAATAT	5280
TATCATGAAT CAAGAATTGG AGTTAATTTA GAATTGTACT TAATTTAGAA TTGTACTTTA	5340
TTAATATTGA GGTAACTTTT TCTTGATAAA GGAAGAAATA ATGGAGAGGA AGTTAGAATG	5400
AAAAAATTCG ACAATTATAT TATTGAGAAG CCTTGCGATT CTAATTCAGA TAAACTGCAA	5460
AAAATCTTAA TAATTGAAAG TTTGGTAGAT GATATTTTGC AATTTTCTCT CAGAATCAAT	5520
AATAGTGTAG GAGAGATTTT CCTCCTACAA CCGTTTTAAA AGAAACTAT CTTTATTCCA	5580
TGTTATTTTG AGGAAGATAT TGTGAAAGTC AAAGATGATG ATAAAGTTGA GTGGAATTTG	5640
TTAGAATTTT AAAAATTTAG AGCATTPTTG GCTTAGTAAT CTGTGTTGAA GGCTCAAAAC	5700
CTATGGTAAA AAAGTAGCTT TGAAAACGTA TTGCCTCCAA AGATTTAGTT AAATAATGAT	5760
TTAACACAAA AAGAAATTAT TGAAGTCTG GAAAGATGTT GTTTCAGTAT TGAGAAAAGG	5820
TGGGAAAAAC TTGCGATTTT CACAGAGAAA GGAAGAAAAA GTATAGAAAT ATAGTCAATT	5880
GAAACAAGAA CAGGATAAAA GAACCTTTTG TGCCATATTT TTCTCCTTTC GCTTTACAAT	5940
TGGATTGAAC ACCTTTATTG TATCGCGTTT GGAGTTTTTT TGGTATAACC TTCGACGCAC	6000
ACCCGCATAG CGGGTGTTTT TTTTGTCTCG CACCTAACGG AGCGAGACAA ACTAATAGTC	6060
ACTTAATCAA AAAACGCACC ATATCAAAAA CTAAAAAGTT TGATATCATG CGTCATGTCT	6120
TAAACTAATT GACTATACTT TCTATTCAAA TGAGCTTTTA ACCAATTGAT TGAGCCAATC	6180
CACTCTTAAA ACCAAAGAGC AATTTCTCGC TTAGCTGACT CTTCTGAATC TGAACCATGT	6240
ACAACATTTT GGATAATCTC ATTTTCTCCA GCAGCTTTTG CAAAATCACC TCGAATAGTG	6300
CCTGGTAAAG CTTCTTCTGG ACGAGTTGCA CCCATCATGG TCCGCCAAGT TTCGATTACT	6360
TTGGGACCAG AAATGACACC CACAAGAACT GGACCTGAAG TCATGAATTC ACGAATCGGT	6420
GGGTAAAAAC TCTGACCAAC CAAGTCCTGA TAGTGCTGGT CAATCAACTC TTCTGAAACC	6480
TGTGAACGAA ACTCCAATTT TTCGATTGTA AATCCACGTT GTTCGATGCG CTTTAACACT	6540
TCACCCACTA GCCCTCTTTT TACACCATCT GGTTTGATGA TAAAGAATGT TTGTTCCATA	6600
CCCGTCTCCT TTGTCAGCTT CTTTCTTTTA TTTTACCACA TTTCTGGAA AAATGGAGAA	6660
AGTTTTCAGA AGAGAGAATG AGAGAACCCT CGGGTCTCTT CATCTCTCTT TATTCTACTG	6720
TTTCTTCCAC AGTTTCAACG GCAGTATCCA CAACTACTTC TGTGTTTTCT TCATTTCTTT	6780
CTTCTCTAC TGGAGGATTA AGGTATTCTT CTTCTGTTGAC AGCATGTGGT TCAAGGTTAC	6840

810

GGTAACGGGC	CATACCAGTA	CCAGCTGGGA	TGATCTTACC	GATGATAACA	TTTTCTTTAA	6900
GTCCAAGGAG	ATGGTCTTTC	TTACCACGGA	TAGCTGCCGC	AGTAAGGACA	CGAGTTGTTT	6960
CCTGGAAGGA	AGCCGCTGAC	AAGAACTGT	TTGTTTCAAG	TGAGGCTTTG	GTAATTCCCA	7020
TAAGGACTGG	GCGACCTGTC	GCTGGAATC	CACCTGCCGAT	AAGGACATCT	TTGTTGGCAT	7080
CTGTAAAGTC	ATTGATATCC	ATGAGGGTAC	CCATGAGAAG	ATCTGTATCA	CCTGGATCCA	7140
TGACACGGAC	TTTACGGATC	ATTTGACGAA	CCATTACCTC	GATGTGTTTG	TCACCGATTT	7200
CTACCCCTTG	GCTACGGTAA	ACTTTTTGTA	CTTCACCGAG	AAGGTACGTT	TCAACTGACA	7260
AGACATCAGG	AACTGCAAGG	AGACGTTTTG	GTTGGATAGA	ACCTTCTGTC	AGAGCAGCAC	7320
CACGCGCTAC	TTGGCCCCCA	ACTTCGACAC	GCATACGAGC	TGTAAATGGA	ACGACATATT	7380
CACCTTCGCC	AGTTTCACCC	TTAACAAAGA	CTTCTTGGT	ACGAGTTGAT	GCATCTTCTT	7440
CGATAGCAGT	AACTTGTCCT	TTAACCTCTG	TAATAACCGC	TTCCCTTTTA	GGATTGCGGG	7500
CTTCAAAGAT	TTCTTGGACA	CGAGGAAGAC	CCTGAGTGAT	ATCGGTATTT	GAGGCAACCC	7560
CACCTGTGTG	GAAGGTACGC	ATTGTAAGCT	GTGTACCAGG	TTCCCCGATA	GATTGGGCAG	7620
CGATTGTACC	AACTGCTTCA	CCAACCTCAA	CCGCATCACC	AGTCGCCAAG	TTGATACCGT	7680
AACAGTGACG	GCAGACACCG	TGACGAGTGT	TACATGTAAA	TACAGAACGG	ATAGTCACTT	7740
CTTCCACACC	AGCATTGACA	ATTTACGCG	CCTTGTCTTC	TGTAATCAAT	TCATTTGGAC	7800
CAATAATCAC	TGCACCAGTT	TCTGGATGTT	TAACAGTTTT	CTTAGTGTA	CGACCGTTGA	7860
GACGCTCTTC	GAGAGACTCG	ATCATCTCTT	TTCTTCTGTC	GATAGAACGG	ATCAAGAGAC	7920
CACGGTCAGT	TCCACAGTCG	TCCTCACGGA	TGATAACGTC	TTGGGCAACG	TCGACCAAAC	7980
GACGAGTCAA	GTAACCTGAG	TCGGCTGTCT	TAAGGGCCGT	ATCGGTCATA	CCTTTACGAG	8040
CACCGTGAGT	TGAGAAGAAC	ATTTCCAATA	CCGACAAACC	TTGCGGAAG	TTTGAAAGGA	8100
TTGGCAATTC	CATGATACGT	CCATTGCGAG	CAGCCATCAG	ACCACGCATA	CCGGCAAGCT	8160
GTGAGAAGTT	TGAGATGTTA	CCACGGGCTC	CAGAGTCCAT	CATCATAACG	ATTGGGTCT	8220
TAGGATCTTG	GTTAGCAATC	AAGCGTTTCT	CAAGTTTTTC	ACGGGCAGCA	CGCCATTACG	8280
CTGTAACAGC	ATTGTAACGC	TCGTCGTCTG	TGATCATACC	ACGACGGAAT	TGTTTGGTGA	8340
TTTGTTGAC	ACGTTTGTGT	GATTCTTCAA	TGATTTCAGC	CTTGTATCA	ACGACTGGGA	8400
TATCGGCAAT	ACCCACTGTC	AATCCTGCAA	GAGTTGAGTG	GTGGTAACCG	AGGTTCTTCA	8460
TGCGGTCAAG	TAGGGCAGAA	GTTTCTGTCT	TACGGAAACG	TTTGAAGATT	TCAGCGATGA	8520
TATTTCCAAG	GTTTTTCTTC	TTGAATGGAG	GTTTGAGCTC	AAGATTGCTG	ATAGCTTCCT	8580
TGATATCTCC	ACCAAGTGGC	AAGAAGTATT	TAGCTGGAAC	ACCTTCTGTC	AAGTTGGCAT	8640

811

TGTTTGGTTC TTGCAAGTAT GGTAGCCCTT CTGGCATGAT ATCGTTGAAG AGAATTTTAC	8700
CAACTGTTGT AAGCAAGACC TTATGTCTTT GCTCTTCTGT CCAAGGCTTG TTGAGGCTGT	8760
CTGTTGCGAT ACCAACACGT GAGTGGAGGT GAACATAACC ATTGCGGTAA GCCATAACCG	8820
CTTCGTCACG GTCTTTGAAG ACCATTCTCT CACCTTCGCG ACCAGCTTCT TCCATGGTCA	8880
AGTAGTAGTT ACCCAAAACC ATGTCCTGAG ATGGAGTAAC TACCGGTTTC CCATCTTTCG	8940
GGTTCAAGAT GTGCTCAGCA GCTAGCATGA GGATACGAGC TTCTGCTTGT GCTTCTTCTG	9000
AAAGTGGTAC GTGGATGGCC ATTTGGTCCC CGTCAAAGTC AGCATTGTAG GCTTCACAGA	9060
CAAGTGGGTG CAAGCGAAGA GCCTTACCAT CAATCAAGAC TGGCTCGAAG GCTTGGATAC	9120
CCAAACGGTG AAGGGTCGGT GCGCGGTTCA AAAGCACTGG GTGTTCTTTA ATCACTTCTT	9180
CAAGGATATC CCAGATACGC TCATCTCCGC GTTCCACCAA GCGTTTAGCT GCTTTGACGT	9240
TTTGCACGAT ATCACGGGCA ACGATTTTAC GCATGACAAA TGGTTTAAAG AGTTCAATCG	9300
CCATTTACAG CGGCACACCA CATTGGTACA TCTTAAGAGT TGGACCAACG GCGATAACTG	9360
AACGTCCTGA GAAGTCAACA CGTTTACCGA GCAAGTTTTG ACGGAAGCGT CCTTGTTTAC	9420
CTTTAAGCAT GTGGCTCAAT GATTTCAATG GACGGCTACC TGGTCTGTG ATTGGACGAC	9480
CACGACGACC ATTGTCAATC AAAGCGTCAA CTGCTCTTG AAGCATACGC TTCTCATTTT	9540
GAACGATGAT ACCTGGTGCA TTAACTCAA GCAAACGAGC CAAACGGTTG TTACGGTTGA	9600
TAACACGGCG GTAAAGGTCA TTCAAGTCAG ATGAGGCAAA ACGGCCACCA TCCAAC TGCA	9660
ACATTGGACG AAGATCTGGT GGGATAACCG GAAGGATGTT AAGAATCATC CATTCAGGTT	9720
TGTTTCCAGA CTTGTAAAAG GCATCCAAAA CATCCAAACG ACGGATGGCT TTGACACGCT	9780
TTTGCTCAGT AGCTGTTTTT AATTCTTCTT TGAGTTCAGC AATTCTTTTT TCAAGATCTA	9840
CTTGCTTCAA AAGGTCTTGG ATGGCTTCCG CACCCATCTT GGCAACAAAT GAACCATAAC	9900
CATATTCACG CAAGCGCTCT CGGTATTTCG GCTCTGTCAT GATAGACTTG TGCTCAAGTG	9960
GTGTATCCTT AGGATCAATC ACCACATAAG CCGCAAAGTA GATAACTTCC TCGAGGGCAC	10020
GAGGGCTCAT ATCAAGGGTC AAGCCCATAC GGCTTGAAT CCCCTTGAAG TACCAGATGT	10080
GAGATACAGG AGCTTTCAAT TCGATATGTC CCATACGCTC ACGACGAACT TTCGTACGCG	10140
TTACTTCAAC CCCACAGCGG TCACAAACAA TTCCTCTGTA ACGAATGCGT TTGTACTTAC	10200
CACAAGCACA TTCCAGTCT TTTGTAGGAC CAAAGATCAC TTCATCAAAG AGTCCTTCAC	10260
GTTCTGGTTT CAAGGTACGA TAATTGATTG TTTCAGGTTT TTTGACTTCT CCATAAGACC	10320
ATGAACGGAC TTTACTTGA GAAGCTAGGG TGATTTCGAT ACTTTTAAAA CGATTACAT	10380

812

CAACCACTAT TTCTCCCTT TCTATTCTAA GTGAAGTCTT TATTCTTGTT CAGCAGCTTC	10440
TTCTGTTGCT TCCGCTTTTG TTGCTTTCTC AGCTTCTTCA GCTTCAAAGG CTGCTTTAGC	10500
CTCTGGGGCT GCTTTTTCGC GGGCTTTTTC AAGGTCATCT ACGTGGATGA CATCTTCGTC	10560
CATTCCCTCA TCCAAGTCGC GAAGTTCCAC TTCTTGGTCA TCTTCGTCTA GGACACGCAT	10620
GTCAAGACCA AGAGATTGCA ATTCTTTGAC AAGAACTCGG AAGGATTCTG GAACACCTGG	10680
TTTTGGAATT GGTTTGCCTT TTGTAATAGC TTCATAGGCT TTCAAACGTC CGTTGATATC	10740
GTCCGACTTG TAAGTCAAGA TTTCTTGAAG GACATTTGAC GCACCGTAGG CTTCAAGAGC	10800
CCAAACCTCC ATCTCACC GAAGTTGTCC ACCAACTGA GCCTTACCTC CGAGTGGTTG	10860
TTGGGTAACA GTTGAGTATG GTCCGACTGA ACGCGCGTGC AATTTATCAT CAACCATGTG	10920
GTGGAGTTTG ATCATGTACA TGAATCCGAC AGAAACACGG TTATCAAACG GTTCACCACT	10980
ACGTCCATCG TAAAGGATCG TTTTGGCATC GCTATCCATA CCTGCTTCTT TAACAGTTGA	11040
CCAAAGATCT TCAGAACTTG CTCCATCAAA GACTGGTGTA GCCATGTGAA TACCAAGAGT	11100
ACGAGCTGCC ATACCAAGGT GAAGCTCCAT AACCTGACCG ATATTCTATC GTGATGGTAC	11160
CCCAAGTGGG TTCAACATGA TGTCGACTGG AGTTCCGTCT GGAAGGTAAG GCATGTCTTC	11220
TACAGGAACG ATACGAGAGA CAACCCCTTT GTTTCCGTGA CGTCCGGCCA TTTTATCTCC	11280
GACCTTAATC TTACGTTTTT GAGCGATGTA AACACGAACC AACATGTTAA CACCTGATTG	11340
CAACTCATCT CCATTTACAC GTGTAAAGAT CTTAACATCA CGAACGACAC CATCGGCACC	11400
GTGTGGTACA CGAAGAGAAG TATCAGCAC TTCACGAGAC TTGTCTCCAA AGATAGCGTG	11460
CAAGAGACGT TCTTCAGCTG AAAGATCTTT CTCACCTTA GGTGTTACTT TACCTACAAG	11520
AATATCACCT TCTTTAACCT CAGCACCAAT ACGGATAATC CCCATTTCTG CAAGGTCTTT	11580
GAGGGCATCT TCACCAACGT TTGGAATTTT GCGAGTGATT TCTTCAGGCC CAAGCTTTGT	11640
ATCGCGCGTT TCTGATTCTG ATTCTTCAAG GTGAACAGAT GTGTAGACAT CGTCTTCAC	11700
CAAGCGTTCT CTCATGATAA CGGCATCTTC GAAGTTGTAA CCTTCCCAAG TCATGTAGGC	11760
AACGATTGGG TTTTGTCCAA GCGCCATTTT TCCATTTTCC ATAGAAGGTC CGTCAGCGAT	11820
GAAATCGCCT TTTTCAACGA CATCACCAAC TTTTACGAGA GTGCGTTGGT TGTAAGCAGT	11880
ACCTGAGTTT GAACGACGGA ATTTTGGAT GTGGTAAACA TCCAATGAAC CATCTTCACG	11940
ACGAATCTCT ACCTGTGAG CATCTGCGTA AGTAACTTTA CCATCATACT GAGCAATCAC	12000
AGCCGCACCA GAATCGTGGG CTGCTTGGTA TTCCATACCA GTACCAACGT AAGGTGCCTG	12060
AGGATTAATC AATGGCACAG CCTGACGTTG CATATTGGCT CCCATGAGGG CACGGTTGGA	12120
GTCATCGTTT TCCAAGAAAG GAATACATGC TGTCGCAACG GCAACTACCT GTTTTGGTGA	12180

813

AACGTCCATG TAGTCAACAA TATTAGCTGG ATACTCTTGG TTGACCCCTT GGTGACGTCC	12240
CATGACAATC TTCTCAGCAA AGGTTCCATC TTCATTGAGA CGAGAGTTAG CCTGAGCTAC	12300
AGTATATTCA TCTTCTTCAT CAGCTGTCAA CCAAACAATT TCGTTCGTGA CAACACCTGT	12360
TTCACGGTCA ACCTTACGGT ATGGTGTGTTG AACAAAACCA TATTTGTTCA AGTGTCCATA	12420
AGATGACAAG TTATTGATCA AACCAGTGT AGGTCCTTCA GGTGTCTCGA TTGGACACAT	12480
ACGACCATAG TGAGTGTAGT GCACGTCACG TACTTCATAT CCAGCACGGT CACGAGTCAA	12540
ACCACCAGGT CCTAAGGCTG ACAAACGGCG TTTGTGAGAC AACTCAGAAA GCGGGTTGTG	12600
TTGGTCCATG AACTGTGACA ACTGTGATGA ACCAAAGAAT TCTTTAACTG CAGCTGTTAC	12660
AGGACGGATA TTGATAATTT GTTGTGGTGT CAAGACTTCA TTGTCCTGAA CAGACATACG	12720
TTCACGGACA TTACGTTCCA TACGAGAAAG TCCCAAACGT ACTTGGTTGG CAAGCAATTC	12780
ACCAACCGCA CGGATACGAC GATTTCGAAG GTGGTCGATA TCATCTACAC GGCCAAGTCC	12840
TTCAGCCAAG TTGAGGAAGT AGCTCATCTC AGCAAGGATA TCTGCAGGAG TCACCGTACG	12900
AACCTTGTC TCTGGGTTAG CATTACCAAT GATCGTTACG ACGCGATCTG GATCAGTTGG	12960
AGCAATAACC TTGAATTTTT GAAGAACAAC AGGCTCAGTC ACAACGGCTG CATCGTTTGG	13020
GATGTAGACA ATCTTGTTCA AGTCGCCATC CAAATGGCTT TCAATGCTTT CAATCACGCT	13080
ACGAGTCATA ATCGTACCAG CTTCTACCAA GATTTCCTCA GTTTCAGGGT CTACCAATGG	13140
CTCTGCAATG GTTTGGTTGA GCAAACGTGT TTTAACATTG AGTTTTTTAT TGATTTTGTA	13200
ACGACCAACT GCTGCCAAGT CATAACGACG TGGGTCAAAG AAGCGAGCTA CAAGCAAGCT	13260
ACGTGAGCTT TCAGCCGTCT TAGGCTCACC TGGACGAAGG CGTTCGTAAA TTTCTTTCAA	13320
GGCTTCGTCT GTACGAGAGT CCATTGGATT CTTGTGGATA TCTTTTCAA CAGTGTGCG	13380
AACCAATTCG CTGTCACCAA AGATATCAAA GATTTCATCA TCACCTGAGA AACCAAGAGC	13440
ACGAACCAAG GTTGTAATG GAATCTTACG AGTACGGTCG ATACGAGTGT AGGTGATATC	13500
TTTGAGTCG CTTTCAAGTT CCAACCAAGC TCCACGGTTA GGGATAACAC TTGAACCATA	13560
GCCCACCTTA CCATTTTTGT CTACTTTGTC GTTAAAGTAA ACACCTGGTG AGCGGACCAA	13620
CTGAGAAACG ATAATACGTT CACCACCATT GATGATGAAA GTACCCATTT CTGTCATGAT	13680
TGGGAAATCA CCAAAGAAAA CTTCTGGGT CTTGATTTG CTTGTTTCTT TATTGATCAA	13740
ACGGAAGGTT AAAAAATTG GTGCTGAGTA GCTAGCATCG TGGATACGAG CTTCTTCTAG	13800
CGTATATTTT GGTTCCTTGA TTTTCATATC AACAAATCC AACTCCATTG TGTCTGTGAA	13860
GTTTGAAATT GGCAATACAT CTTCAAACAC TTCCTTAAGA CCGTGGTCTA GGAAAGCTTT	13920

814

GAATGAGTCA GTTTGAATTT CAATCAAATT TGGTAAGTCA AGAACTTCTT TGATTCTTGA	13980
AAAACCTACGA CGGGTACGAT GTTTCCTCGTA TTGAACGTCA TGTCTGCGCA AGATGATTCT	14040
CCTTTGTAAA TAAGTTCCAA GCCTTGTCAA TCAGGCTTTT CTAATCGTCA TATGGTTGTA	14100
AACCCCTTAT CACCGTGTCC TCTTGACGAA TTTTCAGAAT CTTTAAGCCT CTGTTACAAA	14160
TGCTCAAAAT CTTGAAAAAA AGCACAAAAA GAGCAGCTAA ATCTGACTTT TTCAGAAGAT	14220
TTAACTGCTG TGAGCCTTGT CTGGACAATA TTTCAGACAA AACCTACGAC AAATGATTAC	14280
CCATATTATA CCTATTATAG CTAGATTTTT CAAGGGGTTT CAGTAGGTTT TTGGTAAATT	14340
TTTTCCCATATA GAAAACCTGG CATCACATTC GAATCACGCT ATGGTACAAA AAACCTGAAAA	14400
AACTATTGAC TGAAAATCAT TTTCAAGGTA TAATAATAAA CGTTAAGGCG GTATAGCCAA	14460
GTGGTAAGGC ACGGCTCTGC AAAAGCTTGA TCGTCGGTTC AAATCCGTCT ACCGCCTTCT	14520
ATAACTTGAT TTATCAGGTT TCAAATGAAC AGAAAGCCCA ATTTGAAGGG CTTTTTTTAT	14580
TTTCCCTCGA ATAAATACGT ATAACTTTAA AAACTTTGG AGCGAGTTTG TGGCAGACTT	14640
CTTCCATGG CATAATTCCC TTTTGAAATC AG	14672

(2) INFORMATION FOR SEQ ID NO: 112:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7902 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

AGGAGACTAT TCAAGCCCAA ATTgAGTAGC CCAGCAAAGA CTGTATAGAC TGTGATACGT	60
TTTTCATAGC CATTGGTAAA GAGAATTTGG GAACCAAGAA TGGTATCTAA GGCCAGGATA	120
ATCGTACGAA AAGCGAAGAG AGAGGTCAAG ATGCCGCCCTC CGATATATTT TTCACTACCG	180
TAAAGTAGGA TGGCATTGG TCCTAAAACC ATGAGTCCAA AACTCAGTGG AATGATAAAG	240
AAGTTAAAGA TTCGACTACC TCTATTAACC AGAGAAACAT AGGCTTCTTT GTCTCCTTTC	300
CCCAGATAGT AACTGAGACG AGGCACACTC ACTCCAATTG CACCTGTTAC AACCCAGCT	360
ATAACGGTCA CAATTCGCTG AGCTATGGTA TAGTAACTAA CGTTGACATC AATCCCTGTT	420
TTAACGAGGA AGAGGCGATC TAAAAAAGTG AAGAGCATAT TGGCATTGGC AAAGACTAAC	480
ATGGCTGTCA GAGGGAGAAA GAGTGGTTTA AAATCACTTA GGTGAATTTT AACAAGTTTG	540
ATGTCTCTTT TAATCCAAAA ATAATAATC AGGTAGTTAA TCAGCGTCGA TAAACTCATC	600
ACAAGTGTAT AGACAACAAT ATCGTGTTCA TTTTAAACAA ATAAGAAAAT AGAGACCAGC	660

815

ATCAGGATAC GGATGAAGGC AGTTTTGTAA AAGAGAAAAC TGTAATTTTC CAGAGCTTCA	720
TTGACCCATT CGATTGAAAA AATCTGGGCA ATGAGTTGAA TCCCCATAAC AAGGTAGACC	780
TTTTTGACGA TTGGATTATC AGTAAAGAAG AGAGGATAGG CTAGGATATA GACAGCAGTG	840
GTCAAAATCG TACAAGCGAT GCACAAATAA AAAAGACTAG AAAAGGTTCT GTTAAGATCT	900
TTTTTGTTAT CCTTGACATT ACTGATAGCC CTAAACCGT AGTTATAGAC ACCATAAGTT	960
GCAAAGGGCA AGAAAAATGA CAAAATAGTG TCGACTGAGT TGAAGTAACC ATAGTCAGTT	1020
CGGTCCAAGA CACGCGCGAC ATAGGTTCCA GTTAGGATGG GAAAAATAAT ATTCAAGACA	1080
CGAATTCCCA TGTAAGATAG AGCATTTAAT TTTATACTTT TCATTCAATT TACCTCGTTT	1140
TTCATTATAT CATAAAGTTA GCTAATAAGA AATGAAGGGC AGTAAGTCAA GTAATCACTT	1200
TGAAGTTTCA AATCTTAAGT TTAAAGTTTT CTTAAGGAA AGTATATTAT TCTGAAGGAC	1260
TCTAAAATTT CGCAGCCATT TATTAGTAAT TGCTACAGAA TTCCTAGTCA TTACTAGAAA	1320
TGGACTAGTT TCTTTGAATA ATAGAACTGC ATAATTCTCC TATTCTAGAA GGGGAGGACC	1380
AGTATTTCTT TTATGATAGG ACTAGATTGT GGTATAATAG AGAGAATAAG TTTTTTACTT	1440
AAGACAAAGG AGAAAATAGA TGATTTATGC AGGAATCTT GCCGGTGGAA CTGGCACACG	1500
CATGGGGATC AGTAACTTGC CAAACAATT TTTAGAGCTA GGTGATCGAC CTATTTTGAT	1560
TCATACAATT GAAAAATTTG TCTTGAGGCC AAGTATTGAA AAAATTGTAG TTGGTGTTC A	1620
TGGAGACTGG GTTTCTCATG CAGAAGATCT TGTAAGATAA TATCTTCCTC TTTATAAGGA	1680
ACGTATCATC ATTACAAAGG GTGGTGCTGA CCGCAATACA AGTATTAAGA ACATCATTGA	1740
AGCCATTGAT GCTTATCGTC CGCTTACTCC AGAGGATATC GTTGTTACCC ACCATTCTGT	1800
TCGTCCATTT ATTACACTTC GCATGATTCA GGACAATATC CAACTTGCCC AAAATCATGA	1860
CGCAGTGGAC ACAGTGGTAG AAGCGGTTGA TACTATCGTT GAAAGTACCA ATGGTCAATT	1920
TATTACAGAT ATTCCAAATC GTGCTCACCT TTATCAAGGA CAAACACCTC AAACATTCCG	1980
TTGCAAGGAC TTCATGGACC TTTATGGATC TCTTTCTGAT GAAGAGAAGG AAATCTTGAC	2040
AGATGCATGT AAAATCTTTG TGATCAAAGG AAAAGATGTG GCTTTGGCCA AAGGTGAATA	2100
CTCAAATCTG AAGATTACAA CCGTAACAGA TTTGAAGATT GCAAAAAGTA TGATTGAGAA	2160
AGACTAGTAA AATGATTAAT CAAATTTATC AACTAACTAA GCCTAAGTTT ATCAATGTCA	2220
AATATCAGGA AGAGGCTATT GACCAAGAGA ATCATATCCT TATCCGTCCC AACTACATGG	2280
CTGTCTGTCA TGCGGATCAG CGTTACTATC AGGGAAAACG TGATCCCAAG ATTTTGAATA	2340
AAAAGCTTCC AATGGCAATG ATTCACGAGT CATGTGGAAC CGTCATTCTT GACCCGACCG	2400

816

GAACCTACGA GGTGGTCAA AAAGTTGTCA TGATTCCCAA TCAGTCTCCT ATGCAGAGTG	2460
ATGAAGAATT CTATGAAAAC TACATGACAG GGACCCATTT CTTGTCTAGT GGATTTGATG	2520
GCTTTATGAG AGAGTTTGTT TCTCTCCCTA AAGATCGTGT GGTGGCTTAT GATGCTATTG	2580
AAGATACGGT TGCAGCCATT ACAGAGTTTG TCAGTGTGGG CATGCACGCT ATGAATCGTC	2640
TATTGACTCT TGCTCATAGC AAGCGGGAGC GGATCGCCGT TATTGGAGAT GGAAGTTTAG	2700
CTTTTGTGGT TGCCAATATT ATCAACTATA CTTTGCCAGA AGCAGAGATT GTGGTTATTG	2760
GTCGTCATTG GGAAAAGTTG GAACTCTTCT CATTTGCCAA AGAATGCTAT ATTACGGATA	2820
ATATTCCTGA AGATTTGCC TTTGACCATG CTTTGAATG TTGTGGTGGT GATGGTACTG	2880
GACCAGCTAT TAATGACTTG ATTCGCTACA TTCGTCCTCA GGAACGATT CTCATGATGG	2940
GAGTTAGCGA ATATAAAGTC AATCTCAATA CTCGCGATGC CTTAGAAAAG GGCTTGATTT	3000
TGGTTGGGTC ATCTCGTTCT GGTGCGATTG ATTTTGAAAA TGCTATCCAA ATGATGGAAG	3060
TCAAGAAATT TGCCAATCGT CTTAAAAATA TCCTTTATCT AGAAGAACCT GTAAGAGAAA	3120
TTAAAGATAT TCATCGTGTC TTTGCAACCG ATTTAAACAC AGCCTTTAAA ACAGTGTTTA	3180
AGTGGGAAGT ATAAGTACTG GAGGTTAATT GTGGAGAAAA TCATTAAAGA AAAAATTCT	3240
TCCTTACTTA GTCAAGAAGA GGAAGTCCTC AGTGTGAAC AACTGGGTGG AATGACCAAT	3300
CAAACTATT TGGCCAAAAC AACAAATAAG CAATACATTG TTAAATTCTT TGGTAAAGGG	3360
ACAGAAAAGC TTATCAATCG ACAAGATGAA AAGTACAATC TTGAACTACT AAAGGATTTA	3420
GGCTTAGATG TAAAAAATTA TCTTTTGAT ATTGAAGCTG GTATCAAAGT AAATGAGTAT	3480
ATCGAATCTG CGATTACGCT TGATTCAACG TCAATCAAGA CCAAGTTCGA CAAAATTACT	3540
CCAATATTAC AAATATTCA TACGTCTGCT AAGGAATTAA GAGGAGAATT TGCTCCTTTT	3600
GAAGAAATCA AAAAATACGA ATCCTTGATT GAAGAACAAA TTCCTTATGC CAACTATGAA	3660
TCTGTTAGAA ATGCAGTCTT CTCCTTAGAG AAAAGACTGG CTGACTTAGG TGTGACAGA	3720
AAATCTTGTC ATATCGATTT GGTGCCTGAA AACTTTATCG AATCACCTCA AGGACGACTT	3780
TATTTGATTG ACTGGGAATA TTCATCAATG AATGATCCAA TGTGGGATTT GGCTGCCCTC	3840
TTTTTAGAGT CTGAATTCAC TTCCCAAGAG GAAGAACTT TCTTATCTCA CTATGAGAGT	3900
GACCAAACAC CGGTTTCTCA TGAAAAGATT GCTATTTATA AAATTTTACA AGATACTATT	3960
TGGAGTCTAT GGACTGTCTA TAAGGAAGAG CAAGGTGAAG ATTTTGGTGA CTATGGTGTG	4020
AATCGTTACC AAAGAGCTAT TAAAGGTTTG GCTTCTTATG GAGGTTTACA TGAAAAGTAA	4080
AAACGGAGTT CCTTTTGGCC TTCTCTCAGG TATTTTCTGG GGCTTGGGTC TAACGGTTAG	4140
TGCTTATATC TTTTCGATTT TTACAGATTT GTCACCCCTT GTGGTGGCTG CAACTCATGA	4200



817

TTTTTTGAGC ATCTTTATCT TACTAGCTTT TCTCTTGGTA AAAGAAGGGA AAGTTCGCCT	4260
CTCAATTTTC TTAAATATTC GCAATGTCAG TGTATCATC GGACCCCTGC TAGCAGGCCC	4320
TATCGGTATG CAGGCCAATC TTTATGCAGT TAAGTATATC GGAAGTTCTT TAGCTTCATC	4380
TGTATCGGCT ATTTACCCTG CGATTTCACT TCTATTGGCT TTCTTCTTTT TGAAGCACAA	4440
GATTTTCGAAA AATACTGTAT TTGGGATTGT CTTGATTATT GGAGGGATTA TTGCTCAGAC	4500
CTATAAGGTT GAACAGGTTA ATTCTTTCTA CATTGGGATT CTTTGTGCTT TGGTTTGTGC	4560
TATTGCATGG GGAAGTGAGA GTGTTCTTAG CTCCTTTGCC ATGGAAGTG AATTGAGTGA	4620
AATCGAAGCC CTCCTAATCC GTCAAGTAAC TTCGTTCTTG TCCTATCTTG TGATTGTGCT	4680
CTTCTCTCAT CAGTCATTTA CTGCAGTAGC CAATGGACAA TTGCTAGGTC TCATGATTGT	4740
TTTTGCAGCC TTTGATATGA TTTCTACTT GGCTTATTAT ATCGCTATCA ATCGCTTGCA	4800
ACCAGCCAAG GCTACAGGCT TGAACGTGAG CTATGTAGTA TGGACGGTCT TGTTCGAGT	4860
TGTTTTCTTG GGTGCACCGC TAGATATGCT GACCATTATG ACGTCACTTG TCGTCATTGC	4920
TGGAGTTTAT ATTATTATTA AAGAATAAAG GAGATTCGTG TGAAAGCCAT TATCTTAGCA	4980
GCGGGATTGG GAACTCGCTT GCGTCCTATG ACTGAAAATA CCCCTAAAGC CTTGGTTCAG	5040
GTTAATCAAA AACCTTTGAT TGAGTACCAA ATTGAGTTTC TCAAAGAAAA AGGAATCAAT	5100
GACATCATCA TCATTGTTGG TTATCTTAAA GAACAATTCG ATTACTTGAA AGAGAAATAC	5160
GGTGTTCGTC TCGTTTTCAA TGATAAATAC GCTGACTACA ATAACTTTA CTCTCTCTAT	5220
CTTGTA AAAAG AAGAATTGGC CAACAGCTAT GTTATTGATG CTGACAATTA TCTCTTAAA	5280
AATATGTTCC GCAATGATTT GACACGTTCC ACTTATTTA GTGTTTATCG TGAAGATTGT	5340
ACCAACGAAT GGTTCCTGGT TTATGGAGAT GACTACAAGG TTCAAGACAT TATTGTTGAT	5400
AGCAAGGCAG GTCGCATCCT TAGTGGTGTA TCCTTCTGGG ATGCTCCAAC TGCAGAAAAG	5460
ATTGTCAGCT TTATCGACAA GGCTTATGTA AGTGGTGAAT TTGTTGATCT CTATTGGGAC	5520
AATATGGTTA AGGATAATAT CAAAGAGCTA GATGTCTATG TTGAAGAATT AGAAGGCAAT	5580
AGCATTTATG AGATCGATAG TGTCCAAGAC TATCGTAAAT TAGAAGAAAT TCTTAAAAAC	5640
GAAAAATAAA GATTCCAACA TCTGACAAAA TAGTCGGATG TTTTTCGATT TTTTACGAAC	5700
TTTTACGAAT AGATAGATGA GTAGAAAAAG AAATGGACTT ATTTATGAAA ATCACAAACT	5760
ATGAAATCTA TAAGTAAAA AAATCAGGTT TGACCAATCA ACAGATTTTG AAAGTGCTAG	5820
AATACGGTGA AAATGTTGAT CAGGAGCTTT TGTTGGGTGA TATTGCAGAT ATCTCAGGTT	5880
GCCGTAATCC AGCCGTTTTT ATGGAACGTT ATTTTCAGAT AGACGATGCG CATTGTGCGA	5940

818

AAGAGTTTCA AAAATTTCCA TCTTTCTCTA TTTTAGATGA CTGTTATCCT TGGGATTTGA	5000
GTGAAATATA TGATGCGCCT GTACTTTTAT TTTACAAGGG AAATCTTGAC CTCCTGAAAT	6060
TCCCGAAGGT AGCGGTCGTG GGCAGTCGTG CTTGTAGCAA ACAGGGAGCT AAGTCAGTTG	6120
AAAAAGTCAT TCAAGGCTTG GAAAATGAAC TGGTTATTGT CAGTGGTCTG GCCAAGGGCA	6180
TTGACACAGC AGCTCATATG GCAGCTCTTC AGAATGGCGG AAAAACCATT GCAGTGATTG	6240
GAACAGGACT GGATGTGTTT TATCCTAAAG CCAATAAACG CTTGCAAGAC TACATCGGCA	6300
ATGACCATCT GGTTCCTAAGT GAATATGGAC CTGGTGAACA ACCTCTGAAA TTTTCATTTTC	6360
CTGCCCCGTA TCGCATCATT GCTGGACTTT GTCGTGGTGT GATTGTAGCA GAGGCTAAGA	6420
TGCGTTTCAAG TAGTCTCATT ACGTGTGAGC GAGCAATGGA AGAAGGACGC GATGTCTTTG	6480
CTATTCTTGG TAGCATTTTA GATGGACTAT CAGACGGTTG CCATCATTTG ATTCAAGAAG	6540
GAGCAAAATT GGTCAACAGT GGGCAAGATG TTCTTGCGGA ATTTGAATTT TAAAAATGAC	6600
CTAAGCTAGA ATTCTAAGAA AAAATCAATT TTAAGAGAAA ATGAACCCAA CATTTCCATA	6660
ATAAAACGCA TATTAGCAAG TTTTAAACAC TTGATAATAT GCGTTTTTTC TAAGTGGATT	6720
AGTAGAGTAG AGGATTTTTT TCATATAATA CTCTTCGAAA ATCTCTTCAA ACTACGTCAG	6780
CTTCCATCTG CAACCTCAAA ACAGTATTTT GAGCgaCTtC GTCAGTCTTA TCTACAACCT	6840
CAAAGCAGTG CTTTGAGCAA CCTGTGGCTA GCTTCCTAGT TTGCGCTTTG ATTTTCATTG	6900
AGTATAAGGG AAAGTATAGT GAATTGAAAT AAGATGTGAA CAACTCTATC AGGAAAGTCA	6960
AATTAATTTA TAGAAATATT TTAGCAGCCA AGGTGTACTG TTATAGATTC AATTACACTA	7020
TAATTTAGTG TAATTGAGAA AGGAGAAATG ATTGTGATTG ATGTTGGCTA GGTATGTTC	7080
AATGATTCCT ACCGTCTCAA ATCTTGTCAG TAAGGAAAAA TAAATCTTTC AAAAGTAGAG	7140
ATTACAAGGC TTGTTTAAGA AAGAATTCAA AGACCTTGAC AAATAAAAAAT AAAATGGTTA	7200
TTATAAAAAA TGGTCTGAAA TAGATGATGA TACTTTTCGA AAATCTCTTC AAATACGTCA	7260
GCTCAGCTTT GCCTTGCTGT GTTTGTAGCA AGCTACGGTT AGCTTCCGAG TTTGATTTTC	7320
ATTTACTAGA AATGAAACTG ATGAGAGATA TCAGTAGACA TTTGAGTCAG GATATTATGG	7380
AAAATGATAA AAAGAGCTCG TGAGATTGGC ATATCAGACT ACTAAAGTAT TGAGTTTGTT	7440
AGGATTTTAG CCACTAGTTA GCTGGGAAAG GAAGATATTT GTGACAAATA ATAAACTGTA	7500
TTGTTGATA GAATTTAGAA ATAAATATA TGAAGAATTA GAACTTTCCA GAAGTGATTT	7560
AGCGATTTTA CTATGTGCCA TGCTTATCGC CTCTATCGGA TTAAATATGG ATTCGACTCC	7620
CGTGATTATT GGAGCCATGT TAATCTCTCC TTTGATGACA CCTATTCTGG GAGTGGGGCT	7680
CTCTCTAGCT ATATTTGATT TTAATTTGTT AAGAAAATCT TTTAAATAT TAGCTATTCA	7740

819

AATTCTTGCC AGTCTAATAG CTTCAACACT TTATTTTAT CTTTCTCCCA TTTCGTATGC 7800  
TAGTTCGGAG ATTGTTGCTA GAACCTCTCC GACTATTTGG GATGTTCTCA TTGCTTTTGT 7860  
AGGAGGGATA GCAGGTATCA TTGGTGCTAG GAAAAAAGAG AC 7902

(2) INFORMATION FOR SEQ ID NO: 113:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18627 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

GAAGTTGAAA TGGCCAGCTG ATGAGCAATA TCGGTCATAG AAATCTTCTC AATCAACTTT 60  
TGCGCAATTT TTTGGTTGAT AATACGAGGA ATTTGGTGAT TTTTCTTGAC GATAGAAGTT 120  
TCAGCGACCA TCATTTTGA ACAGTGATAG CACTTGAAAC GACGCTTCT AAGTAGAATT 180  
CTAGTAGGCA TACCAGTTGT CTCAAGGTAA GGAATCTTAG ACGGTTTTTG AAAGTCATAT 240  
TTCTTCAATT GGTTCGCA CTCAGGGCAA GATGGGGCGT CGTAGTCCAG TTTGGCGATG 300  
ATTCCTTGT GTGTATCTT ATTGATGATG TCTAAATCT GGATATTAGG GTCTTTAATG 360  
TCTAGTAATT TTGTGATAAA ATGTAATTGT TCCATATGAA TCTTTCTAAT GAGTTGTTTG 420  
GTCGCTTTC ATTATAGGTC ATATGGGACT TTTTTCTAC AATAAAATAG GCTCCATAAT 480  
ATCTATAAGG GATTACCCA CTACAAATAT TATAGAGCCA AAAATCCTTT GTTTACTAAA 540  
CAAGGGATT TTCTTTTGTC TCTGCTCCTT TTTTGATATA ATAGTTCTAT GTTAAATCA 600  
GAAAAACAAT CACGTTATCA AATGTAAAT GAAGAATTGT CCTTCCTATT GGAAGGCGAA 660  
ACCAATGTTT TGGCTAATCT TTCCAACGCC AGTGCTCTCA TAAATCACG TTTTCCTAAT 720  
ACCGTATTG CAGGCTTTTA TTTGTTGAT GGAAGGAAT TGGTTTAGG CCCCTTCCAA 780  
GGAGGTGTTT CCTGCATCCG TATTGCACTA GGCAAGGGTG TTTGTGGTGA GGCAGCTCAC 840  
TTTCAGGAAA CTGTTATTGT TGGAGATGTG ACGACCTATC TCAACTATAT TTCTTGAT 900  
AGTCTAGCTA AAAGTGAAAT TGTGGTGCCG ATGATGAAGA ATGGTCAGTT ACTTGGAGTT 960  
CTGGATCTGG ATTCTTCAGA GATTGAGGAT TACGATGCTA TGGATCCAGA TTATTTGGAA 1020  
CAATTTGTCG CTATTTTGCT TGAAAAGACA GCATGGGACT TTACGATGTT TGAGGAAAAA 1080  
TCTTAATGTA TCAAGCACTT TATCGAAAAT ATAGAAGTCA AAATCTCTCC CAGTTAGTTG 1140  
GTCAAGAAGT TGTGGCTAAG ACTCTTAAAC AAGCGGTGGA GCAAGAGAAA ATAAGTCACG 1200

820

CTTATCTTTT TTCTGGTCCT CGTGGAACGG GAAAAACCAG TGTTCCTAAA ATCTTTGCCA	1260
AGGCTATGAA CTGTCCCAAT CAAGTGGGTG GCGAACCTTG CAATAACTGC TATATTTGTC	1320
AAGCAGTGAC GGACGGTAGT TTAGAAGATG TCATTGAAAT GGATGCAGCT TCTAATAATG	1380
GGGTAGATGA AATTCGCCAA ATTCTGTGATA AATCTACCTA TGCGCCTAGC CTTGCTCGTT	1440
ATAAGGTTTA TATCATAGAT GAGGTTCACTA TGCTGTCTAC AGGGGCTTTT AATGCCCTCC	1500
TAAAGACGCT GGAAGAACCA ACACAGAATG TAGTCTTTAT TTTGGCCACT ACTGAATTGC	1560
ACAAGATTCC TGCTACTATT CTATCCCGTG TGCAACGTTT TGAGTTTAAA TCAATTAAGA	1620
CACAGGATAT TAAGGAACAT ATTCACTATA TCTTAGAAAA AGAAAATATC AGTTCTGAAC	1680
CAGAGGCTGT GGAAATCATT GCCAGACGGG CGGAAGGTGG AATGCGGGAC GCCTTGCTTA	1740
TTTTGGATCA AGCCCTGAGT TTGACACAGG GAAATGAGCT GACGACTGCT ATCTCTGAAG	1800
AAATTACTGG CACCATTAGC CTATCAGCCT TGGATGATTA TGTGGCGGCC TTGTCTCAAC	1860
AGGATGTTCC CAAAGCTTTG TCTTGCTTGA ATCTTCTTTT TGACAATGGT AAGAGCATGA	1920
CTCGTTTTGT GACCGATCTT TTGCACTATT TAAGAGACTT GTTAATTGTT CAAACAGGGG	1980
GAGCAAATAC TCATCATAGT TCAGTCTTTG TAGAAAAATT GGCACCTTCCT CAAAAAATC	2040
TGTTTGAAAT GATTGCTTA GCAACAGTGA GTTTAGCAGA TATTAAGTCT AGTTTGCAAC	2100
CCAAGATTTA TGCTGAAATG ATGACCGTCC GTTTGGCGGA AATCAAGTCC GAACCAGCTC	2160
TATCAGGAGC GGTGAAAAAT GAAATGCTA CGCTGAGACA GGAAGTTGCC CGTCTCAAAC	2220
AAGAGCTTTC TAATGTAGGT GCGGTTCCCTA AACAAGTTGC ACCAGCTCCT AGTCGACCAG	2280
CTACGGGCAA AACAGTCTAT CGTGTGATC GCAATAAAGT GCAATCTATC TTACAAGAGG	2340
CCGTCGAAAA TCCTGATTTA GCACGTCAAA ATTTAATTCT TTTGCAGAAT GCCTGGGGAG	2400
AGGTAATTGA AAGTCTAGGT GGGCCGGACA AGGCTCTGCT AGTTGGTTCT CAACCGGTTG	2460
CTGCCAATGA ACACCATGCT ATTCTTGCTT TTGAGTCTAA CTTCAATGCT GGTCAAACCTA	2520
TGAAACGAGA CAATCTCAAT ACCATGTTTG GTAATATCCT CAGTCAGGCG GCAGGTTTTT	2580
CACCTGAGAT TTTAGCTATT TCCATGGAGG AATGGAAAGA AGTTCGCGCA GCCTTTTCAG	2640
CCAAAGCCAA ATCTTCTCAA ACTGAAAAAG AAGTAGAAGA AAGCCTGATT CCAGAAGGAT	2700
TTGAATTTTT GGCTGATAAA GTGAAGGTAG AGGAAGACTA AAGAAAGATT TCATGATACA	2760
ATAAGTTTAT GAATAACAA CAATTTATTA TTATGGCGCT GTTTACAGCT GCTGAGACCT	2820
ATTTTTTCAA TGAAGCCTGG ATGACTGGCC GCTATATTAT GGCAGCCTTT TGGGCAATTT	2880
TACTCTTTAG AAATTTCCGA GTCAGTTATG TGATGGGCAA AATCGTTGAT GTCATCGATC	2940
AGCATTTTAA TAGGAAAGAC TAGCCCTCAG CTTCCAGACA AAATCAAAGC CTTTTAGGCT	3000

821

TTTTTTTGT ATACTAGAAA AGTATATTTA TAGAATTTTT GCTCTATTTT TGGGGAAATC	3060
AGACGTTTTT CTAGTAAGTA CTGTAAAAGT TTTGAAAAAG AAAGGAACTA TCATGTCAGT	3120
ATTAGAGATC AAAGATCTTC ACGTTGAGAT TGAAGGAAAA GAAATTTTAA AAGGGGTAA	3180
CCTGACCCTG AAAACAGGAG AAATTGCCGC TATCATGGGA CCAAATGGTA CAGGTAAATC	3240
GACTCTTTCT GCCGCTATCA TGGGAAATCC AAACATGAA GTAACATAAG GTGAAGTTTT	3300
GTGATGAGC GTAAACATCC TTGAGTTGGA AGTGGATGAG CGTGCGCGTA TGGGACTTTT	3360
CCTTGCTATG CAATACCCAT CAGAAATCCC TGAATTACC AATGCTGAGT TTCTTCGTGC	3420
CGCTATGAAT GCGGGTAAAG AAGATGATGA GAAGATTTCA GTTCGTGAGT TTATTACTAA	3480
GCTAGATGAA AAAATGGAAT TGCTCAACAT GAAAGAAGAA ATGGCAGAGC GTTACCTCAA	3540
CGAAGGCTTC TCTGGTGGTG AGAAAAACG CAATGAAATT CTTCAACTTT TGATGTTGGA	3600
GCCAACATTT GCTCTTTTGG ACGAGATTGA CTCAGGTCTT GATATTGACG CTCTTAAAGT	3660
TGTGTCTAAA GGTGTCAATG CCATGCGTGG TGAAGGTTTT GGTGCTATGA TCATCACTCA	3720
CTACCAACGT CTTTTGAACT ATATCACACC TGATGTGGTA CACGTGATGA TGAAGGTGCG	3780
TGTTGTCCTT TCTGGTGGTC CAGAAATGGC TGCAGGTTTG GAACGTGAAG GATACGCAAA	3840
ATTAGCTGAA GAACTTGGCT ACGACTACAA GGAAGAATTG TAATCCCTC GTATCTTTTA	3900
GGAGAAGTAA ATGACTAGAG AAAATATTAA ACTTTTTTCA GAAATGCACG CTGAACCAAG	3960
CTGGTTGGCT GATCTCCGTC AAAAGCTTTT TGACAAGATT GAGACTTTGG AATTACCAGT	4020
TATTGAGTGT GTCAAATTC ACCGTTGGAA TCTGGGTGAT GGAACGATTA CAGAAAATGA	4080
GCCATCAGCA AATGTTCCAG ATTTACAGC TTTAGATCAT CACTTGAAGT TGGTGCAAGT	4140
AGGAACTCAA ACTGTTTTCG AACAACTCC AGTTGAGTTA GCTGAACAGG GTGTTGTCTT	4200
CACAGACTTT CACTCAGCTT TAGAAGAAAT TCCAGAGCTG ATCGAAGAAT TCTTCATGTC	4260
ATCTGTTAAG TATGATGATG ACAAGTTGGC GGCTTACCAC ACAGCTTACT TTAACAGTGG	4320
TGCTGTACTC TATATTCCAG ATAACGTAGA AATCACAGAG CCAATTGAAG GAATTTTCTA	4380
CCAAGATAGC GATAGCAATG TGCCGTTTAA CAAGCATATT ATGATTATCG TTGGTAAAAA	4440
TTCTAAGATT AGTTATCTGG AGCGTTTAGA GTCACGCGGT GAAGGAAGTG ACAAAGCAAC	4500
TGCCAATATC ACAGTGGAAG TGATTGCACG TTCTGGTGGC CAAGTCAAGT TTGCTGCTAT	4560
CGACCGTCTA GGTGAAAACG TCACTGCCTA CATTAGCCGT CGTGGTAAAT TAGGCAACGA	4620
TGCAAGTATT GACTGGGCTA TCGGTGTCAT GAACGAAGGA AATGTCGTTG CTGATTTTGA	4680
TAGTGAATG ATTGGTAATG GTAGCCATGC TGACCTCAAG GTTGTAGCTC TTTCAAGTGG	4740

822

TCGTCAGGTA CAAGGGATTG ATACTCGTGT AACTAACTAT GGCTGCAACT CAATCGGAAA	4800
CATTCTACAA CATGGGGTTA TCCTTGAAAA AGCAACTTTG ACTTTCAATG GTATCGGCCA	4860
CATCATCAAG GGTGCTAAGG GAGCAGATGC GCAACAAGAG AGCCGTGTTT TCATGCTTTC	4920
AGACCAAGCG CGTTCAGATG CTAACCCAAT TCTTTTGATT GATGAAAATG ACGTAACTGC	4980
AGGCCATGCA GCCTCTATTG GTCAGGTAGA TCCAGAAGAT ATGTACTACC TCATGAGTCG	5040
TGGCTTGGAT AAGGCAACTG CAGAGCGTTT GGTGTTCGT GGTTCCTTG GATCTGTTAT	5100
CGTGGAGATT CCAGTCAAGG AAGTTCGTGA TGAATGATT GCAACTATCG AAGAGAAATT	5160
GTCAAAACGC TAAGGGGCGAG CCTATGTTAG ATGTAGAAGC GATTGCGAAG GATTTTCCAA	5220
TTTGTAGATCA GATTGTCAAT GATGAACCTC TGGTCTATCT GGACAATGCT GCGACCACAC	5280
AAAAACCACT AGTAGTTCAG AAAGCTATTA ACAGCTACTA TGAGCAGGAC AATGCCAATG	5340
TTCAACCGTG TGTCCATACC TTAGCGGAAC GAGCGACAGC TTCTTATGAA GCTGCTCGTG	5400
AAACCATTCG TAAGTTTATT AATGCAGGCT CTACAAAGGA AGTTCCTCTT ACCAGAGGAA	5460
CGACAACCAG CCTTAACTGG GTGGCAGCT TTGCTGAGGA AATTCTCACT GAGGGAGACC	5520
AGGTCTTGAT TTCAGTAATG GAACACCATT CTAATATCAT TCCATGGCAG GAAGCTTGTC	5580
GAAAGACTGG AGCAGAGCTT GTCTATGTCT ATCTTAAAGA CGGTGCCTTG GATATGGAGG	5640
ATTTGCGAGC TAAATTGACT GATAAGGTTA AATTTGTTTC CCTAGCTCAT GCCTCCAATG	5700
TTCTTGGTGT GGTCAATCCG ATCAAGGAAA TCACTCAATT AGCCACCAA GTTGGGGCAA	5760
TTATGGTAGT GGATGGTGCT CAATCTACAC CTCATATGAA GATTGATGTC CAGGACTTGG	5820
ATCTGGACTT TTTGCGCTTT TCGGGTCACA AGATGGCTGG TCCGACTGGT ATCGGTGTCC	5880
TTTACGGCAA AGAAAAGTAT CTTGAGCAAA TGTCTCCAGT AGAATTGGC GCGGAGATGA	5940
TTGATTTTGT CTACGAGCAA TTTGCTAGTT GGAAGGAATT GCCTTGAAA TTTGAGGCTG	6000
GAACGCCAAA TATGGCAGGA GCTATTGGAC TTGCGACTGC AGTTGATTAT CTGGAAAAGA	6060
TTGGTATGGA TGCCGTTGAA GTCATGAAC AGGAATTGAT TGCGTACGTC TATCCAAAAC	6120
TGCAGGCAAT TGAGGGATTG ACCATTTACG GTTCTCAGGA TTTGGCTCAA CGTTCGGGTG	6180
TTATTGCCTT TAACCTAGGT GATCTCCATC CTCACGATCT TGCGACGGCT CTGGATTATG	6240
AAGGAGTGGC TGTTCTGCTT GGTCAACATT GTGCGCAACC CTTGCTTCAG TATTGGAAG	6300
TCCCAGCAAC AGCTCGTGCA AGTTTTTATA TCTACAATAC CAAGGCAGAT TGCGACAAAAC	6360
TAGTCGATGC CCTACAAAAG ACAAAGGAGT TTTTCAATGG CACTTTCTAA ACTAGATAGC	6420
CTTTATATGG CAGTGGTAGC AGACCATTCG AAAAATCCAC ATCACCAAGG GAAGTTAGAA	6480
GATGCTGAGC AAATCAGTCT CAACAATCCG ACTTGTGGGG ATGTCATCAA CCTCTCTGTG	6540

823

AAGTTTGATG CAGAGGACCG TTTGGAAGAT ATTGCTTTTC TAAATTCAGG ATGCACGATT	6600
TCAACTGCTT CTGCTAGTAT GATGACAGAT GCCGTTT TAG GAAAAACCA ACAAGAAATT	6660
TTAGAACTGG CGACTATTTT TTCTGAAATG GTTCAAGGGC AAAAAGATGA GCGTCAAGAC	6720
CAACTTGGAG ACGCGGCATT CTTGTCAGGT GTTGCCAAAT TCCCTCAAAG AATCAAGTGT	6780
GCAACCCTAG CTTGGAATGC CCTTAAGAAA ACAATTGAAA ATCAAGAAAA ACAGTAAGAC	6840
AAGTTTCTTT TGTCTTATGA ATTATTAGAA ATGAAGAAAG AAAGGATACT ATGGCTGAAG	6900
AAAGAGTAGA ACCAAAACCA ATTGACCTTG GTGAATATAA ATTTGGTTTC CATGACGATG	6960
TAGAGCCTGT CTTATCGACA GAAAAGGAC TCAACGAAGG TGTATTTCGT GAATTATCTG	7020
CTGCTAAGGG TGAGCCTGAG TGGATGTTGG AGTTCGTTT GAAGCTTAT GAAACCTTCA	7080
AAAAATGCC CATGCAAACT TGGGGAGCAG ACTTGTGAGA GATTGACTTT GATGACTTAA	7140
TCTACTACCA AAAACCATCT GACAAACCAG CCCGTTCTTG GGATGATGTA CCTGAAAAGA	7200
TTAAAGAAAC CTTTGAACGT ATCGGGATT CAGAAGCTGA ACGTGCTTAT TTAGCAGGGG	7260
CTTCTGCCCA GTACGAGTCA GAAGTGTTT ACCACAACAT GAAGGAAGAG TTCCAAAAAT	7320
TAGGTATTAT CTTTACAGAT ACAGATTCCG CACTCAAGGA ATACCCAGAC TTATTTAAAC	7380
AATACTTTGC GAAGTTGGTA CCGCCGACAG ATAACAAGTT GGCAGCCCTC AACTCAGCAG	7440
TATGGTCGGG TGGAACTTTT ATCTACGTGC CAAAAGGTGT CAAGGTAGAT ATTCCACTTC	7500
AACTTATTT CCGTATCAAT AACGAAAATA TAGGTCAGTT CGAACGTACC TTGATTATCG	7560
TTGATGAGGG AGCAAGCGTC TACTACGTAG AAGGATGTAC AGCACCAACA TATTCAAGCA	7620
ATAGCTTACA CGCTGCCATT GTAGAAATTT TTGCTTTGGA CGGAGCTTAT ATGCGTTATA	7680
CAACTATCCA AAAGTGTCT GATAACGTCT ATAAGTTGGT AACAAAGCGT GCTAAGGCTC	7740
AAAAGGATGC CACTGTTGAG TGGATTGATG GAACTTGGG TGCCAAAACG ACTATGAAAT	7800
ATCCATCTGT TTACCTTGAT GGAGAAGGAG CGCGTGGTAC CATGCTCTCT ATCGCCTTTG	7860
CTAATGCAGG GCAACACCAA GACACGGGTG CTAAGATGAT TCACAATGCT CCACATACCA	7920
GCTCGTCTAT TGTGTCTAAA TCCATCGCTA AAGGTGGAGG AAAGGTTGAC TACCGTGGAC	7980
AAGTCACCTT TAACAAGAAC TCTAAGAAAT CTGTTTCCCA CATTGAATGT GATACCATTA	8040
TCATGGATGA CTTGTCAGCA TCAGATACTA TTCCATTAA TGAAATTCAC AACTCGCAAG	8100
TGGCTTTGGA ACACGAAGCC AAAGTATCTA AGATTTGAGA AGAGCAATTG TATTATCTCA	8160
TGAGCCGTGG ATTGTCAGAA TCTGAGGCAA CTGAAATGAT TGTCTGGGA TTTGTAGAAC	8220
CCTTTACAAA AGAACTTCCA ATGGAATACG CAGTTGAGCT GAACCGCTTG ATTAGCTATG	8280

824

AAATGGAGGG ATCAGTTGGA TAAAATTGGA TTTTATACTC TTCGAAAATC TCTTCAAACC	8340
ACGTCAGCAT CGCCTTACCG TATGTATGGT TwCTGAlTCG TCAGTTTCAT CTACAACCTC	8400
AAAACAGTGT TTTGAGCAAC tGCGGCTAGC TTCCTAGTTT GTTCTTTGAT TTTGAGTATT	8460
AGATTTACTC AAAATCAAGG ATTTTGAAGA TGAACCTGTA TCAAAAAATC GCGGTTTAAA	8520
ATCGCGATTT TTTATAATTT CTCGTTAACA AAGCGGACAA ACTGATTCCA CCAAACCTTT	8580
AAGAAGAAGG CTTTTTCAAT TTTCTTGCTT GCTACCATTT CGAAACTAGG GCGTCTGTG	8640
GTGATGTAAC CTTGACCAAT CAAGTCCTTG TCTTCATAAG TCAAAATGGCC AACCCTGTT	8700
CCAGCTTCAA GTGGTGCTGG GATTGCTTTG GAATCAGGTG TGAATTGAAC AGATTGGGAA	8760
GATTGATTCC CAACACGTTT GATTAGATAG ATATCCTCTG GAGCCACTGC AGTTACTGTA	8820
TCTTCTTTTC CATCTGTAC AGGGGCTTG CTATCTTGAT AGGCATCGCC TTGTGAACG	8880
ATTTTGCGAA GTGTAAATGT AGAAGAAATA TAATCCATTA GGAAGATGT AGCTGTAAAT	8940
CGAGCGTAAG GATTATTGTC TTGATGATCT GCATTTAAAA CAACTGTGAT GACTCTCATG	9000
CCTTTTTCGA CAGTAGTACC AACAAAAGAC TCTCCAGCCT TATCTGTTGT TCCTGTTTTT	9060
AGCCCATCAA AACCACCACG GTAAGCAGGC ATACCTTCTA ACATGTAGTT GGTGAAGTG	9120
ATTGTCATCC CAGCAAAAGT AGAAGAAGGT TTTTGGTGA TTTCTAAGAC TTGTGGGTAT	9180
TTTTTGATGA GGTGCGAGC AACGATAGCG ACATCATAAG CACTAAGCTT ATTTTCCTCA	9240
TCTTTTTTAG AACCTGGGTA AATGTTATCC CCTAGAGTTT CATTGTTAAG ACCTGTCGTA	9300
TTGACAACAG TGGCATCCTG AATTCCCCAT TCCAAGAGTT TTGCCCGCAT CATATCGACG	9360
AAATCTTTTT CTGAGCCAGC AATTTTCTCA GCTAGGGCAA TAGCGGCGCT GTTGGCACTA	9420
GATACCAGAG TTGCTTCAAG CAACTCTTCG ACAGTATAAT TACGGGCCTC CATAGGAATA	9480
TTACTGGCTT CAGAATTTGT CGTCAATTGA TAAGGATAAT CAGAAATATC TACAGGAGTG	9540
GAGAGGGTAA TACTTCCGTT TTCCAAAGCT TCATAGACCA GATAAACAGT AATCAATTTT	9600
GTTATGGAAG CAATTTGAC AGGTTGCGTT GCATCCTTCT CATAGAGAAT TTTACCAGTA	9660
TTTGCCCTCAA CAGCAATCGC ATGTTTAGCG GCAATGGTAA AATCTTGAGC AACAGCAGTA	9720
GAAGCACCCC CTAAAAGAGA GACAGTTAAC AAAGTTAAAA ATATTTTTTT CATAGTAGTC	9780
TTATTCTATC ATAAAGAAAA AAAATATTCT TGCTTTAATA ATTCATCTGT TAAGCTTTTT	9840
GAAAATATGG TAAAATAAAG TAAGGGAGGT AACTCATGTT TCGTAGAAAT AAATTATTTT	9900
TTTGACCAC AGAAATTTTA CTCTTAACCA TCATCTTTTA CCTATGGAGA CAGATGGGGT	9960
CTTTGATTAA CCCTTTTGTT AGCGTGCTTA ATACAATTAT GATTCCATTT TTATTAGGGG	10020
GCTTTTTTTA TTATTTGACA AACCCTATTG TTACTTTCTT AAATAAAGTC TGTAAACTCA	10080



825

ATCGTTTGCT TGGTATTTTA ATTACCTTGT GTACTTTGGT CTGGGGAATG GTCATAGGTG	10140
TTGTCTATCT CTTACCTATT TTGATTAATC AGTTATCTAG TTTGATTATA TCTAGTCAAA	10200
CTATTTATAG TCGAGTACAA GACTTAATCA TAGACTTATC TAATTATCCT GCGCTCCAGA	10260
ATTTGGATGT AGAAGCTACA ATTCAGCAGT TAAACTTATC CTATGTTGAT ATTCTTCAAA	10320
ATATCCTAAA TAGCGTATCA AATAGTGTGG GGAGCGTCTT GTCAGCTCTT ATCAGTACTG	10380
TTTTGATTTT GATTATGACT CCAGTTTTTT TGGTTTATTT CTTATTAGAT GGACATAAAT	10440
TCTTGCCCAT GCTTGAAAGA ACGATTCTAA AGAGGGATCG CTTGCATATT GCAGGCTTAT	10500
TAAAGAATTT AAATGCGACG ATTGCTCGCT ATATTAGTGG AGTTTCGATT GACGCAATCA	10560
TTATAGGTTG TTTGGCTTAT ATTGGCTATA GTATTATTGG TTTAAAATAT GCTTTAGTTT	10620
TTGCCATTTT TTCTGGTGTA GCCAATTTAA TTCCTTATGT GGGGCCAAGT ATTGGTTTGA	10680
TTCCCTATGAT CATCGCAAAT ATATTCACGT ATCCCCATAG ACTGCTGATT GCAGTGATTT	10740
ATATGCTTGT TGTTCAGCAG GTAGATGGCA ATATCTTATA TCCTCGAATC GTAGGAAGTG	10800
TTATGAAGGT TCATCCAATC ACGATTTTAG TTTTACTTTT GTTGTCAGC AATATCTATG	10860
GTGTAGTTGG AATGATTGTC GCAGTGCCAA CCTATTCTAT CTGAAAGAA ATTTCTAAGT	10920
TCTTATCCCA TTTGTATGAA AATCATAAAA TAATGAAAGA ACGAGAAAGA GAATTAGCTA	10980
AGTAAAAGTC AGGAGAACCC TGATTTTTCT TTAAGTGAAG TGGCCTTTAG ATTAGAAGAC	11040
TGAAAATAAG TTAAAGTCTT AAATAATTT TCACAGCTAA GAATAGTAGA AGTTAATCTG	11100
ATAAAAATCG AAAAAACCAG TGAATTCTG TGTCAGGGTA AGTTCCACTG GTTTTCATAG	11160
TCTATTAAAG TTCGAATGAA ACCTATTTAT AGTAGATTGA AACTAGAATA GTACACCTCT	11220
AATTCTAAAA CATGTTTGA AATCGATTTG ACTGTCCTGA TCTATTCGTT CTATTCTTAT	11280
TTCAATTTTAC TATATTTTGG TGCAATAAGT GAAAAGTAGT CCGAATAATA TAAGGATTGA	11340
TTTTATAGTT TTTAACTCA AATGAATTGA AATAAGAGA GTACGAAAAT TCTCATCTGA	11400
AAGTATTTTA GAATAATTCT CTTGCTGAAT TTCTTCAAAA CAGATAGCTT CATCTTAGGT	11460
ATGTGATTTT TTTTGCATT TTTGAGTTAG ATAAGGTATA ATGATTTTAT TGTCTTTTGG	11520
GGTCGTTACG GATTCGACAG GCATTATGAG GCATATTTTG CGACTCGTGT GCGACGTAA	11580
ACGCTCAGTT AAATATAACT GCAAAAAATA ACACCTCTTA CGCTCTAGCT GCCTAAAAAC	11640
CAGCAGGCGT GACCCGATTT GGATTGCTCG TGTTCATGA CAGGTCTTAT TATTAGCGAG	11700
ATACGATTAA GCCTTGCTA GCGGTTTGAT AAGAGATTGA TAGACTCGCA GTTTCTAGAC	11760
TTGAGTTATG TGTCGAGGGG CTGTTAAAAA AATACATAAC CTATGGTTGT AGACAAATAT	11820

826

GTTGGCAGGT	GTTTGGACGT	GGGTTGACT	CCCACCGGCT	CCATTATTCC	TTTGCATTCT	11880
TTTGCATTCC	TTGGTAAAAC	GTTGTAAAT	CAACGTTTTT	TATTTTATC	TTTGGTATTC	11940
CTTTGCATTC	TTTGTCTAAA	AAGGGAGTCA	CAAACAGACC	CTATTTTAAA	AAAGGATAGA	12000
AAAAAGGATA	CAACATTGT	CGCATCCTAA	AAATAATCTT	TTTTCGACGG	AAGACATGGG	12060
ATTCGAACCC	ACGCACGCTA	TTACACGCCT	ACCGCGTTTC	CAACACGGCC	TCTTAAGCCT	12120
CTTGAGTAAT	CTTCCAATAC	TTACTCAAAT	AGTCTACCAT	AAAGGCTCTT	ATCTTGCAAT	12180
AAAAATTCTA	GAAATAAGAA	AAATGATAGA	TTTTGAAAGA	AAATGATAAA	AAATGCTTGA	12240
CTTCGAAAGA	AAGTATGATA	GAATGAATAG	TGTAAACGAT	AACAGGAGGT	GATTCACTGT	12300
TAAAAACAGA	ACGTAAACAA	CTAATTTTAG	AGGAGTTAAA	TCAACATCAT	GTAGTTTCTC	12360
TAGAAAAATT	AGTTAGTTTG	CTAGAAACGT	CAGAATCAAC	GGTTCGAAGA	GACTTGGATG	12420
AGTTGGAAGC	GGAAAACAAG	CTTCGTCGTG	TGCATGGTGG	AGCAGAACTC	CCCTACTCCT	12480
TACAGGAAGA	AGAAACCATT	CAAGAAAAAT	CTGTCAAAAA	CCTTCAAGAA	AAGAAATTGC	12540
TGGCTCAGAA	AGCAGCCTCT	CTCATTAAAG	AAAAAGATGT	CATCTTTATC	GATGCTGGAA	12600
CAACAACCTGC	TTTTTTGATT	CATGAATTGG	TCAATAAGAA	TGTTACAGTT	GTGACCAACT	12660
CCATTCACCA	TGCCGCTCAG	TTGGTTGAAA	AGCAGAACTC	AACTGTCATG	GTTGGAGGAA	12720
ACGTCAAGAC	GGCGACAGAT	GCTAGTATCG	GGGGCGTTGC	TCTTAACCAG	ATTAACCAAT	12780
TGCACTTTGA	CCGTGCCTTT	ATCGGAATAA	ATGGTGTGTA	CGATGGCTAT	TATACGACTC	12840
CTGATATGGA	GGAGGGAGCT	GTGAAAAGAG	CTATTTTGGA	GAATGCCAAG	CAGACCTACG	12900
TCTTGGTGGA	TTCGTCAAAA	ATTGGACAAA	CTTGCTTTGC	CAAGGTAGCC	CCACTCAAAAC	12960
GCGCTATCGT	TATCACTAGT	CAAGGGCATG	AGCTCTTGCA	GGTTATTAAG	GAGAAAACGG	13020
AGGTAATAGA	AGTATGATTT	ATACAGTCAC	ACTCAATCCA	TCCATTGACT	ATATCGTTTCG	13080
TTTGGACCAA	GTCAAAGTTG	GTAGTGTCAA	TCGTATGGAC	AGTGATGATA	AGTTTGCTGG	13140
TGGGAAAGGA	ATCAATGTCA	GCCGTGTCTT	GAAACGTTTG	AATATACCAA	ATACAGCGAC	13200
GGGATTTATC	GGTGGCTTTA	CTGGTAAATT	TATCACAGAT	ACTTTAGCAG	ACGAAGAAAT	13260
CGAGACACGT	TTTGTCCAGG	TGGCAGAAGA	TACTCGTATC	AATGTTAAAA	TCAAAGCAGA	13320
CCAAGAAACA	GAAATCAACG	GAACGGGTCC	AACTGTTGAA	TCGGTTCAGC	TAGAAGAATT	13380
GAAAGCTATT	TTATCTAGTC	TGACAGCACA	AGATACACTT	GTCTTTGCAG	GTTCAAGTGC	13440
TAAAAATCTA	GGCAATGTTA	TCTATAAGGA	TTTGATTTCC	TTGACGCGCC	AGACTGGTGC	13500
GCAAGTGGTC	TGTGACTTTG	AAGGACAGAC	CTTAATTGAT	AGTTTGGACT	ACCAGCCTCT	13560
TCTTGTAATA	CCAAACAATC	ATGAACCTGG	AGCGATTTTT	GGGGTTAAAC	TCGAAAGTTT	13620

827

AGATGAAATT GAGAAATACG CTCGTGAGTT ACTGGCTAAG GGTGCTCAAA ATGTTATTAT	13680
CTCTATGGCT GGTGATGGTG CCCTTCTTGT CACATCTGAG GGAGCTTACT TCGCTAAACC	13740
AATCAAAGGA ACAGTCAAAA ATTCAGTTGG AGCTGGTGAT TCTATGGTTG CTGGATTAC	13800
AGGTGAATTT GTCAAATCAA AAGACGTAGT AGAAGCCTTC AAATGGGGAG TGGCTTGCGG	13860
AACGGCAACT ACCTTCTCAG ATGACTTGGC AACGGCGGAA TTTATTAAAG AAACATATGG	13920
AAAAGTTGAG GTAGAAAAAC GATGAAAATT CAAGACCTAT TGAGAAAAGA TGTGATGTTG	13980
CTAGATTTGC AGGCAACTGA AAAACAGCT GTCATCGACG AGATGATTAA AAATTTGACA	14040
GACCACGGTT ATGTAACAGA TTTTGAAACA TTAAAGAAG GAATTTTGGC GCGTGAAGCT	14100
TTGACTTCTA CTGGTTTGGG TGATGGAATC GCAATGCCTC ACAGCAAAAA CGCTGCTGTC	14160
AAAGAAGCGA CAGTTCTATT TGCTAAGTCA AATAAGGGTG TTGACTACGA GAGCTTGGAT	14220
GGACAAGCAA CTGACCTCTT CTTGATGATT GCAGCTCCAG AAGGTGCCAA TGATACTCAC	14280
TTGGCAGCCT TGGCAGAATT GTCTCAATAC TTGATGAAAG ACGGTTTTGC AGACAAACTT	14340
CGTCAAGCAA CATCTGCAGA CCAAGTTATC GAACTTTTTG ACCAAGCTTC AGAAAAAACT	14400
GAGGAACTTG TTCAAGCACC TGCTAATGAC TCTGGTGAAT TTATCGTAGC TGTACAGCT	14460
TGTACAACAG GTATTGCCCA CACTTACATG GCCCAAGAAG CCCTTCAAAA AGTAGCTGCT	14520
GAAATGGGGG TTGGTATCAA GGTGGAACC AACGGTGCTA GCGGTGTTGG AAATCAACTA	14580
ACTGCAGAAG ATATCCGTAA GGCTAAAGCT ATTATCATTG CAGCAGACAA GGCCGTGAA	14640
ATGGATCGAT TTGATGGAAG ACCATTGATC AATCGTCCAG TTGCTGACGG TATCCGTAAG	14700
ACAGAAGAGC TAATTAACCT GGCTCTTTCA GGAGATACTG AAGTCTACCG TGCCGCTAAT	14760
GGTGCCAAAG CTGCAACAGC CTCTAACGAA AAACAAAGCC TTGGTGGTGC CTTGTACAAA	14820
CACTTGATGA GTGGTGTATC TCAAATGTTA CCATTGCTTA TCGGTGGTGG TATCATGATT	14880
GCCCTTGCCT TCTTGATTGA CGGTGCTTTG GGTGTTCCAA ATGAAAACCT TGGCAATCTT	14940
GGTCTTACC ATGAGTTAGC TTCTATGTTT ATGAAAATTG GTGGAGCTGC CTTGGTMTG	15000
ATGCTTCCAG TCTTTGCGGG TTATGTTGCC TACTCTATTG CTGAAAAACC GGGTTTGGTA	15060
GCAGGTTTCG TGGCTGGTGC TATTGCCAAA GAAGGTTTTC CCTTTGGTAA AATTCCTTAT	15120
GCCGCAGGTG GTGAAGCAAC TTCAACTCTT GCAGGTGTCT CATCTGGTTT CCTAGGTGCC	15180
CTTGTGGTG GATTTATCGC AGGTGCCTTG GTTCTTGCCA TCAAGAAATA CGTTAAAGTT	15240
CCTCGTTCAC TCGAAGGTGC TAAATCAATC CTTCTATTGC CACTTCTTGG AACAATCTTG	15300
ACAGGATTTG TTATGCTAGC TGTGAATATC CCAATGGCTG CAATCAACAC TGCTATGAAT	15360

828

GACTTCCTAG GCGGTCTTGG AGGAGGTTCA GCTGTCCTTC TTGGTATCGT CCTTGGTGGA	15420
ATGATGGCTG TTGACATGGG TGGACCAGTT AATAAAGCAG CTTATGTCTT TGGTACAGGT	15480
ACGCTTGCGAG CAACTGTTTC TTCAGGTGGT TCTGTAGCCA TGGCAGCAGT TATGGCTGGA	15540
GGAATGGTGC CACCACTTGC AATCTTTGTC GCAACTCTTC TTTTCAAAGA TAAATTTACT	15600
AAGGAAGAAC GTAACCTCTGG TTTGACAAAC ATCATCATGG GCTTGTTCATT TATCACTGAG	15660
GGAGCGATTG CATTTGGTGC CGCTGACCCA GCTCGTGCGA TTCCAAGCTT CATCCTTGGT	15720
TCAGCAGTAG CAGGTGGACT CGTTGGTCTT ACTGGTATCA AACTCATGGC GCCACACGGA	15780
GGAATCTTCG TTATCGCCCT TACTTCAAAT GCTCTCCTTT ACCTCGTTTC TGTCTTGGTA	15840
GGAGCAATCG TAAGTGGTGT GGTATTATGGT TACCTACGCA AACCACAAGC ATAAAAAATA	15900
GAAAAATGAA AAGATTGGAC CGTTTGGTGC AGTCTTTTTC TCTTCCCGAA ATGCCTGTGA	15960
AATATGGTAT AATAGAAGAA TGGCAAACAA GAATACAAGT ACAACAAGAC GGAGACCGTC	16020
TAAAGCAGAA CTGGAAGAA AAGAAGCGAT TCAACGAATG TTGATTTCGT TAGGAATTGC	16080
GATTTTATTG ATTTTCGCAG CCTTCAAATT AGGGGCTGCA GGTATAACCC TTTATAATTT	16140
AATTCGCTTG CTAGTGGGTA GCCTAGCTTA TCTGGCGATA TTCGGCCTAT TAATCTATCT	16200
CTTCTTTTTC AAGTGGATAC GAAAACAGGA AGGACTCTTA TCTGGCTTTT TCACCATATT	16260
TGCTGGCTTA CTCTTGATTT TTGAGGCCTA CTTGGTTTGG AAATATGGTT TGGACAAGTC	16320
CGTTCTAAAA GGGACCATGG CTCAGGTTGT GACAGATCTG ACTGGTTTTC GAACGACTAG	16380
CTTTGCTGGA GGGGGCTTGA TCGGGGTCGC TCTTTATATT CCAACAGCCT TTCTCTTTTC	16440
AAATATCGGA ACTTACTTTA TTGGTTCTAT CTTGATTTTA GTGGGTTCTC TCCTAGTCAG	16500
CCCTTGGTCT GTTTACGATA TTGCTGAATT TTTCAGTAGA GGCTTTGCCA AATGGTGGGA	16560
AGGGCACGAG CGTCGAAAAG AGGAACGCTT TGTCAAACAA GAAGAAAAAG CTCGCCAAAA	16620
GGCTGAGAAA GAGGCTAGAT TAGAACAAGA AGAGACTGAA AAAGCCTTAC TCGATTTGCC	16680
TCCTGTTGAT ATGGAACGG GTGAAATTCT GACAGAGGAA GCTGTTCAAA ATCTTCCACC	16740
TATTCCAGAA GAAAAGTGGG TGGAACCAGA AATCATCCTG CCTCAAGCTG AACTTAAATT	16800
CCCTGAACAG GAAGATGACT CAGATGACGA AGATGTTTCTG GTCGATTTTT CAGCCAAAGA	16860
AGCCCTTGAA TACAACTTC CAAGCTTACA ACTCTTTGCA CCAGATAAAC CAAAAGATCA	16920
GTCTAAAGAG AAGAAAATTG TCAGAGAAAA TATCAAAATC TTAGAAGCAA CCTTTGCTAG	16980
CTTTGGTATT AAGGTAACAG TTGAACGGGC CGAAATTGGG CCATCAGTGA CCAAGTATGA	17040
AGTCAAGCCG GCTGTTGGTG TAAGGGTCAA CCGCATTTCC AATCTATCAG ATGACCTCGC	17100
TCTAGCCTTG GCTGCCAAAG ATGTCCGGAT TGAAGCACCA ATCCCTGGGA AATCCCTAAT	17160

829

CGGAATTGAA GTGCCCAACT CCGATATTGC CACTGTATCT TTCCGAGAAC TATGGGAACA 17220  
 ATCGCAAACG AAAGCAGAAA ATTTCTTGGA AATTCCTTTA GGGAAGGCTG TTAATGGAAC 17230  
 CGCAAGAGCT TTTGACCTTT CTAAATGCC CCACTTGCTA GTTGCAAGGT CAACGGGTTC 17340  
 AGGGAAGTCA GTAGCAGTTA ACGGCATTAT TGCTAGCATT CTCATGAAGG CGAGACCAGA 17400  
 TCAAGTTAAA TTTATGATGG TCGATCCCAA GATGGTTGAG TTATCTGTTT ACAATGATAT 17460  
 TCECCACCTC TTGATTCCAG TCGTGACCAA TCCACGCAAA GCCAGCAAGG CTCTGCAAAA 17520  
 GGTGTGGAT GAAATGGAAA ACCGTTATGA ACTCTTTGCC AAGGTGGGAG TTCGGAATAT 17580  
 TGCAGGTTTT AATGCCAAGG TAGAAGAGTT CAATCCCAG TCTGAGTACA AGCAAATTCC 17640  
 GCTACCATTG ATTGTCGTGA TTGTGGATGA GTTGGCTGAC CTCATGATGG TGGCCAGCAA 17700  
 GGAAGTGGAA GATGCTATCA TCCGTCTTGG GCAGAAGGCG CGTGCTGCAG GTATCCACAT 17760  
 GATTCTTGCA ACTCAGCGTC CATCTGTTGA TGTCATCTCT GGT'TTGATTA AGGCCAATGT 17820  
 TCCATCTCGT GTAGCATTTG CGGTTTCATC AGGAACAGAC TCCCGTACGA TTTTGGATGA 17880  
 AAATGGAGCA GAAAAACTTC TTGGTCGAGG AGACATGCTC TTTAAACCGA TTGATGAAAA 17940  
 TCATCCAGTT CGTCTCCAAG GCTCCTTTAT CTCGGATGAC GATGTTGAGC GCATTGTGAA 18000  
 CTTCAATCAAG ACTCAGGCAG ATGCAGACTA CGATGAGAGT TTTGATCCAG GTGAGGTTTT 18060  
 TGAAAATGAA GGAGAATTTT CGGATGGAGA TGCTGGTGGT GATCCGCTTT TTGAAGAAGC 18120  
 TAAGTCTTTG GTTATCGAAA CACAGAAAGC CAGTGCGTCT ATGATTTCAGC GTCGTTTATC 18180  
 AGTTGGATTT AACCCTGCGA CCCGTCTCAT GGAAGAACTG GAGATAGCAG GTGTCATCGG 18240  
 TCCAGCTGAA GGTACCAAAC CTCGAAAAGT GTTACAACAA TAAAAAATA GCTTCTTTCC 18300  
 AAGTTTGGAG GGAAGCTATT TTAGTGGCTA TTGATTGCTT TTATTTTCTG AAGTTGGCGC 18360  
 ATTGGACTGT TTTTCGTTTT CAGTAGCAGG TTTACTTGAA GCAGGAGTAG AAGAGTCCCTG 18420  
 AGTTGCTGTT TTCTGATCTT CTTTTTCTC TTCCTTGACG CTAGATTTTG GTGTTTCTC 18480  
 TTGCTGTGTT TTTCTTGAC TAGTGTAGT CTCTTAGTT GGACTGGTGT TTTCTTAGG 18540  
 GGATTCCTTT TGGATTTCTT TGACAATGGT TGTCGTCTGG CTTGTCGTAG GTTCTTTTTT 18600  
 AATATTTTGT TTATTATCCA AGGCGTT 18627

(2) INFORMATION FOR SEQ ID NO: 114:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2560 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

830

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

TAAAATACGT TACCTTGCTT CTGCACGTTT AGCAGGTAAG TCATTGAAAT TTAAAGATCA	60
AGATATTACA ATTGAAGAAA CGACTGAAAC AGCTTTTGAA GGAGTTGATA TTGCTCTCTT	120
TTCAGCAGGT AGTTCTACAT CAGCTAAGTA TGCACCATAC GCAGTAAAAG CTGGCGTGGT	180
AGTAGTAGAT AATACATCTT ATTTCCGTCA AAATCCAGAT GTTCCTTTGG TTGTTCCAGA	240
GGTCAATGCT CATGCACTTG ATGCTCACAA CGGAATCATT GCCTGCCCTA ATTGTTCAAC	300
AATTCAAATG ATGGTGGCTC TTGAGCCGGT TCGCCAAAAA TGGGGCTTGG ACCGTATCAT	360
TGTTTCAACT TATCAAGCCG TTTCAGGTGC TGGTATGGGA GCAATTCTTG AGACACAACG	420
TGAACCTCGT GAAGTCTTGA ATGATGGTGT GAAACCACGT GATTTGCATG CGGAAATCTT	480
GCCTTCAGGT GGTGACAAGA AACATTATCC TATCGCCTTT AACGCTCTTC CACAAATTGA	540
TGTTTTCACT GATAATGATT ACACGTACGA AGAGATGAAG ATGACCAAGG AAACCTAAGAA	600
AATTATGGAA GATGATAGCA TTGCAGTATC TGCAACATGT GTGCGTATTC CAGTCTTGTC	660
AGCTCACTCT GAGTCTGTTT ATATCGAAAC AAAAGAAGTG GCTCCAATCG AAGAAGTAAA	720
AGCAGCTATC GCAGCCTTCC CAGGTGCTGT TCTTGAAGAT GATGTAGCTC ATCAAATCTA	780
TCCTCAAGCT ATCAATGCAG TTGGTTCGCG TGATACCTTT GTTGGTCGTA TCCGTAAAGA	840
CTTGATGCA GAAAAAGGAA TTCACATGTG GGTGTTTCA GATAACCTTC TCAAAGGTGC	900
TGCTTGGAAC TCAGTTCAGA TTGCTGAAAC TCTTCATGAA CGTGGATTGG TTCGTCCAAC	960
AGCCGAATTG AAATTTGAAT TAAAATAGTC ATATCGTTTA GGAGTTCAGA TGAACCTCTT	1020
CTTTGAAATA GAGAGGTGTT TTCGTGTCTT ATCAAGATTT AAAAAATGT AAAATCATT	1080
CAGCCTTTAT TACCCCTTTC CATGAGGATG GTTCCATTAA CTTTGATGCT ATTCCAGCCT	1140
TGATTGAGCA TTTATTGGCC CATCATACGG ATGGAATTCT TCTCGCAGGA ACGACTGCTG	1200
AGAGTCCAAC TTTGACCCAC GATGAGGAGT TGGAGTTGTT TCGGCTGTA CAAAAGGTT	1260
TCAATGGACG CGTTCCTTTG ATTGCGGGTG TAGGTACTAA TGATACGCGT GACTCTATTG	1320
AGTTTGTCAA AGAAGTAGCG GAATTTGGTG GTTTCGCAGC TGGGCTTGCT ATTGTTCTCT	1380
ACTACAACAA ACCTTCTCAA GAAGGGATGT ATCAGCACTT TAAGACTATT GCAGATGCTT	1440
CTGACCTACC AATTATTATC TATAACATTC CAGGGCGTGT ACTTGTCGAA TTGACTCCAG	1500
AAACCATGCT TCGCTTGGCT GACCATCCAA ATATTATCGG TGTCAAAGAA TGTACTAGCT	1560
TGGCTAATAT GGCTTACTTG ATTGAGCACA AGCCTGAAGA GTTCTTGATT TATACAGGTG	1620
AGGATGGAGA TGCTTTCCAT GCCATGAACC TTGGGGCGGA TGGGGTTATT TCTGTTGCCT	1680

831

CTCATACAAA TGGGGATGAA ATGCACGAGA TGTTTACTGC GATTGCAGAA AGCGATATGA	1740
AGAAAGCCGC AGCAATTCAG CGTAAATTCA TTCCTAAGGT TAATGCTCTC TTCTCTTATC	1800
CAAGTCCTGC TCCAGTTAAG GCAATTCTTA ACTATATGGG ATTTGAAGCT GGACCCACTC	1860
GTCTACCTCT TGTTCAGCA CCAGAAGAAG ATGCCAAACG CATTATCAAG GTTGTCTAG	1920
ATGGCGACTA CGAAGCAACT AAGGCAACTG TAACAGGGGT CTTAAGACCA GATTACTAAT	1980
AAAGACAATA AAATCCGGCT CTTTGTCAAC TGTAGTGGGT TGAAGTCAGC TAAGCTCGAG	2040
AAAGGACAAA TTTTGTCTT TCTTTTTTGA TATTCAGAGC GATAAAAATC CGTTTTTTGA	2100
AGTTTTCAAA GTTCCGAAAA CCAAAGGCAT TCGCCTTGAT AAGTTTGATG AGATTATTGG	2160
TCGCTTCCAA TTTGGCGTTT GAATAGGGTA GTTGAAGGGT GTTGACGATT TTCTTTTTGT	2220
CCTTTAGAAA GGTTTTAAAG ACAGTCTGAA AAATAGGATG AACCTGCTTC AGATTGTCCT	2280
CAATGAGTCC GAAAAATTTC TCCGGTTCCT TATTCTGAAA GTGAAACAGC AAGAGTTGAT	2340
AGAGCTGATA GTGATGTTTC AAGTTTGTG AATAGCTCAA AAGCTTGTTT AAAATCTCTT	2400
TATTGGTTAA GTGCATCGA AAAGTAGGAC GATAAAATCG CTTATCACTC ACTTTACGGC	2460
TATCCTGTTG AATGAGTTTC CAGTAGCGCT TGATAGCCTT GTATTGCGGA TTTTCGATGA	2520
AACTGATTCA TGATTTGGAC ACGCACACGA CTCATAGCAC	2560

## (2) INFORMATION FOR SEQ ID NO: 115:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11303 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

TATTGGATTT CCCTTGCAAT CAGTTTATGG GACAAGCACC CGGCAGCGCA GAGGAAATCA	60
ACGCCTTCTG TAGCCTACAT TTTCAAACCA CCTTCCACG TTTTGCCAAG ATTAAGGTCA	120
ACGGTAAGGA AGCAGACCCT CTCTATGTCT GGTACAAGA CCAGAAATCC GGCCCACTAG	180
GAAAACGAGT CGAATGGAAT TTCGCTAAGT TTCTCATCGG TCGAGATGGG CAACTCTTTG	240
AACGCTTTTC TTCAAAAACA GACCCAAAAC AAATTGAAGA GGCGATACAA ACTCTACTAT	300
AATTCACAAT CTCACTATGA TTAGGTTTCC TTAAACCTGA TGAATAGTGA GATTTTTTGA	360
TGGGCTTTGA CTTAAATAGA AAAACACCCC ATGATATGAA ACATGAAGTG TTGTAAAGTC	420
TATGTTGTAG GTGCTTATTT CACAATTTC AATGTGACCAG TGATAACGAA TACCATACAG	480

832

AATCTTCATA TACACTAAAC AAATGACTTT CTAATTATTT CAATTAGTTT TGGCTAGTAA	540
ATATCATTTT CAACAAACGC CCTCTCAATT CCTTATCTTG ATGATGCAAG ATATTCATTA	600
AGTCATGAGA GTTTTTTCGCA TTGATGAATT GATTTAACAA TCTATCTTTT AATTCATATG	660
GAAGAGAAGC TGTCTTTAGT AGTCTAAAAA CTTTCGTCATT TAAAGATGTC CTTTTATTAT	720
CTTTCCATTG AAATTTAGCT GTATCATTTT TATTTGGCAA TTCAATTATA GACACATTCG	780
TTCTTTTAAA ATGAATTCTA TGTTTTCTAT TGCTTGGAAC GATACTAGAA TCTCCTTGTA	840
ATGCTAACTC TACCATTCCC ATTTCCCAAT CGATTGATAA TCTTGTTTTA TATCTTTGAC	900
CATTTTGATC TTCAAGCATT TCAAAAGAAT GTTGTTTTCC TGGGAATACA TACCAATCTA	960
CAACTTCAGG TAAATCAACA CCCATACCTA TCTCAGAACC AACCAAGGGA ATGATTGCAC	1020
CACCTTTTGC AAACACAGGC GTAGTCGAGA TGTCCCTATA AACACTTAAC TTCACACCAC	1080
CTGTGTATTT TTTCTCTGAA AAGAAGTCAT ACCATTCAAC TTCAGGGAAC CATACATCTA	1140
CTTTTGACAG TTGGAATGTC AAATCCATCT TTTCTACAAT GGGAGCCACC ATCAGTTCTG	1200
TTCCAAAAAA GTATTGGTTT GGAACATTAT AGCTCTCATC ATTCTCTGGA TAGAAATAAT	1260
AGATTGGACT GATTAATGGG GCACCTTCCT CATGTGTCCTG TACATTATG GTATATAGAT	1320
AGGGAATCAT CTGATGTCTC AAACGAAGGT ATTTCTTCAT AATCTTAGAT GTTGTTCCTG	1380
AAAAAAACCA AGGTCTTTTA CTATTAAAAG GACTTCTAGA ACTATGTAAT CGAGTAATCG	1440
GACTAAAAAC ACCAAACTGT AGCCATCTAG TTTGTAGCTC TTCGTCATAA TCCCCAACA	1500
TATGTCCACC GATATCATGA CTCCACCAAC TATAACCGAT ATTAGATGCT GTCGCTGTAA	1560
AATAGGGTTG AAATCTTAAG GAATTCACAC TAATAATAGT ATCCCCTGAA AAACCAACAG	1620
GGTAGCGGTG ACTACCAGGA CCTGCATATC TTGATAAAAT CAAACCACCT TCTGCATTTT	1680
TACAACTATC CTGATAGTGA TAATGGTTTA AAAGCCAAAG TGGATCTAGC ATACCTTGTC	1740
TCCCTTGTTG CCAGTCAATC CACCAAAAAT CTACTCCCTG CTTTCTAGT TCATAATGAA	1800
CATCTTTAAA GTAGGCTTCC CTAAGAGAGG GATTAAAAAA ATCAAAAATA GCAGGTTCTT	1860
CTAGTTCTAC ATTTAACCCC AACCGTTTTC CGATTGAGG ATAAGCTTCT TCATAAGCCC	1920
GTATCCCATC AGCAGGATGG ACATTTAAGG AGAGTTTTCAG CTTTCTATCA TGAAGTTGTT	1980
GCAATAACTG TTCTGGATTT GGTATTAAGT TTCTATTCCA ACTATATCCT GTCCAGCCAC	2040
TTCCAAAGCG AGCTGGAATG TCAGTTATAT GCCAATCCAT ATCTAACACA CCGATAGATA	2100
ATGGAATTTT CTCTGTTTCA AATCTGTCTA TTAAATCCAA GTATTATCC GACGTATAAG	2160
GCCAATATCT ACTCCACCAA TTGCCTAAAG CATATCTTGG CAACAAGGGT GTTGAACCAG	2220
TCAAATGGTA AAAATCTCTG ATTGCTCCTC TATAATCATG CCCATAGGCA AAGAAATACA	2280



833

GGTCAATTTG ATTTTCTCTC TCAATATAAC CAGATTGTTT ATCCCAAATA AATCCTTGAG	2340
AATCATCCAA TAAGGCTATA CCATTTCTGGC TAATAATTCC ATCTTCTAAC GAGATTGCTC	2400
CATCTGCCTT ATCCAGAGTC CGAGCTGTTT CTTTAAACGT TTCAATAGAT TCACCAAAAT	2460
ACCAGCGACT ACCATATACG GCAAAATTTC CTTTAAATTC TATAAATAAA TTTTCGGCGT	2520
TAAATTCCTC TTTATTAAAG TGCAGATGAA AATAGTCCGT CATAATATCT AGTACGTTTG	2580
ATGTCTCGAT ATAATCTAAC GAAATTTGGC CAAATCTCT ATTATAGATA AGTTGTGTCG	2640
TTCTATCCTC AAAACTTCCA GTTTGAGAGT ATTCTAACCT TACTAGCTTG TCTGTTAATA	2700
CAGAGATTCG ATAAACTCT CCCTTAAAAA TTTCAATTT GTTTTCCTCC TTTTATGGTA	2760
GCATAAAAC AGAACGCACC ATTTTGTATG CGTTTTCAT TATTCTGAAT GCAATGTCT	2820
ATCTGTTATA TCTATGACAA ATAATAGTCA ATTGAAAAA TGCAGTGGAC AAAATATCTT	2880
TTAACAAACC AAGAGTTTAT TAAAGAGTTA TCACTTTTCA ACTTTTCTAA GCTTATGCAG	2940
TTGTGAAACA AACTACTTTT AAATATTAA CTAAGATAGG ATTGATAAAT AATTTCAAAC	3000
TCTTACTAGC AATCATACGA TATTCAAGCT CACGTGCTTT TTTCTTCTT GCTTATTTCT	3060
TAGAACTGAA GAACCCGGAT CGGTATATAA ATTATCCGGA TCAACATAGT CATAAGATTC	3120
ATAACAGTTG CGCTTCATTA AGTCATCCCC AGAGCAAGAG CTTCATCTCG TAATTTTTC	3180
ACATCACTAA CCGTAGGTCG CCATCCTTCA ATCATATTTG TACTTAAAGC ATACCAAACA	3240
CTCTTAAAAA CGGATCGGTT TTCAAAAGCT ATTCCCATGA TTGTCATCTT TTCTTTATCT	3300
ATATCTAAGG ACATATGCTA CCTCCTTTAG ATACATTATA CCATGTTTCT CTGTAGCTTT	3360
TAAAAATTTT ATTTTGTGTT TCATATCTAA GTTTTCAGCA CGCTTATCCT ATTTTATAAG	3420
CCTCAAACCC AAATATAAAA CGCATTCTTT TTGCTTTTTT ACTATTGTAT CGTATTCTAC	3480
GATAACATAC TTTACTTTAT TGTTTTTTTA AATAACAGCA GTTCCCTGTT TATCAACTAT	3540
TCGAACTACT TTCTATTTTG CTTCATACCC TACATAGCGA AAAAATATGA AAAAGCAGAG	3600
AAGAATATCT TAAAAAGACC TCTTCACTGC TAATATTAAC ACTCATTATT TAACTATAT	3660
GGATTCTATC ATCGAGTATA CTTTTTACT TATTAGATAC CTTGCTCTTC TTTCACCAAT	3720
TTTTGATCAT ATACACGGAT GAATGGAAGA TAGACTAGGA ATGCTGCAAA TGCACATACT	3780
AGAGCAACTA ATACAGCTCG AAGATCTGCT GTCCCTAAGA AAGCTCCAAT CCCTACTGGA	3840
GTTGGCCATG GAACCTGTGC GATAATTGGC TTAATAAAGT TTAGAGAATT CGCTACGTAA	3900
TAAATAGTAG CAGTAACCAT TGGTGCTAAA ATAAATGGTA TAGCCAAGGC TGGATTATAG	3960
ATAATAGGTA ATCCAAAAAT TAATGGTTCA TTAATATTAA ATAAGGCTGG AACTACAGAT	4020

834

GCTCGTCTTA	TTGCTTTAAG	CTGTTTCAGAT	TTAGAGGCAA	AAGCAATATA	TAAACATAGT	4080
CCTAAAGTTG	CACCAGAACC	ACCTGCAATT	ACAAACATAT	TAGAAAATTC	ACCTGCAACA	4140
GCGAAGTGCC	CGCCAGCAGC	ATTTTCAGCC	ATGTTAGCAA	GAGCAATTGG	ACTAACAAAT	4200
GCAAAAACAA	TGTTTCGCACC	GTGGATACCT	ACAATCCAAA	GTAGTTGAGT	CAATAGATAA	4260
ATAATCATT	AACCAATCCA	CGAATTAGTC	AGATTGGATA	CAATACCAA	TGGAATTGCA	4320
ATGACTTTAA	AAATATCTGT	TCCCATTGCT	ACAAGAAGAC	CGTTGATAAA	GATAACAACA	4380
AATGCAACAA	CAAATCCCGG	AACCAAAGCG	GTAATCCAC	GAGAACTCC	TTCTGGAACA	4440
GCTTCAGGCA	TTTTAATAAC	CCAATTATGT	TTAACACACA	TACGATAAAT	AAGAACAGTC	4500
ACAATTGCCA	TAATGATTGC	GGTAAAAATC	CCTGTTGTCC	CAAAACGTGC	GAATACATTT	4560
CCCATTGCCC	ATCCATCTGC	AATTACTGCA	CCTTCTTTTA	GACTTGTCAC	AGTCTTCATC	4620
ATTCCACCAT	CAAAAATGAT	TTGCGGTACT	GTCATGACAA	AAGCCATCAA	GGCAAGCAAG	4680
GCACCATTAA	GAGGATTCAT	ATTGAGTTCT	TCTTCTCTG	CATAAATTTT	TGTCAATTCA	4740
TATGCAAGTG	ATAGAACGAA	ATAAAGAGAT	AGAGAACCCA	TAGTCGCATA	GTTTGCAACC	4800
ATGTAAAGTG	ATGTGAATTT	ATCAAATGAA	GCAGAGAAAA	TATCTGCCAC	AATTGGCCAA	4860
AATGAGAAAG	CTTGTGGCAA	AATACTGAAT	ACCAAAAACA	TTGATCCTAC	AATAGTAAAT	4920
GGTACAGCAG	CCATACCTGC	AGCCGTGATA	GCACGTACTA	CTTTAAACTG	AGCAAGTTTG	4980
CCCATTGGTC	CCATAACATG	GTTTTCAAGA	AAACCAAACA	ACCCGTTTTG	TTGATCCATA	5040
AATAGACCTC	CTTAATAAAA	CATAATAATT	TTTACTTTCT	AAAGACTAGT	TTCAAATACA	5100
AATTATACTA	GATCAGGATT	ATAAACTAAG	TGAGTTCTTT	TCCAATTGGA	CAAATTGTTG	5160
ATAAGCCTTA	TCTGTTCGTT	TATAAATTTT	TTTAATTCTT	CTAATGTCTA	ACAAACTCAG	5220
AACTAAACCT	AATAGAAGAA	CTACAAAAAC	AAATAAACGT	GCTACTTGGT	TATTTTCAAA	5280
AATCGGAAAA	AGATTCTTAA	ACCAACTTGT	CCAAGTTAAA	ACAAGTAATC	CTATTGAAAT	5340
AAGCATTGT	ATTCTAACAA	ACATTAGTGT	TATCCCAAC	TTTCTTTTCC	TATTTCCATA	5400
AAGTTTAAAT	TGTTCAACAG	TTGCTAAAAT	AGAAAATACT	ATGAGCATAA	TGGGGAAAAT	5460
AATAATAGGC	GAGGGACTAA	TAACTGACT	CAAAAGCCAA	TAAATATTCC	CAAAAAAGAA	5520
GAGTGCTATT	GAATAACGTA	GAAGAAGATA	TCGATTGAAA	AAAGTATTAG	TTAGAGCCAT	5580
CTCTCGACGT	TGTTGTTCAA	TCTTTTGTG	TTCTTTTTTA	TCCATATCAT	TTCTCTCTTA	5640
TATAACAACA	CATATTTAGT	TAACTTTCTT	ATAAAGAGCT	AACATTTCTT	TTGCTACTTC	5700
TAATAATGTC	ATAGTGGTCA	TTAAATGATC	TTGAGCATGT	ACCATGATAA	TTTCAATTTT	5760
AATTTCCACT	CCACTTGCGT	ATTCTTGCAA	GAGTTTGTTT	TGTGCATGAT	GCGCTTCAAG	5820

835

AATTATCTCA TTTGATTGAT TTAATTTACT TTCTGCATCA TCAAACTAC CTTCTCTCAT	5880
TTTTGCAAAT GCTTCATGTA TTTCTGACCT TGCATTTCCC GAATGCAGGA TAATTTCAAA	5940
TGCTGCAACC TGCAGTTCCT CTTGATTCAT ATAAACCTCC TATTTTATCT TCTCAAATAT	6000
GTAAATAAAA TCTTCAAAGT TATTGCAAGA TATTAGCTGA TTTTGCAATT CATCATCTC	6060
TGTCAGAGAG ACTATCTTTT TAGTCACAGT TGCCAAACCT TCGTTCCCAT ATATTGATGG	6120
AGATAGAAGA AATACTAGCT GGACATGTGA ACTTTGATTA TCCCAGAGTA ACGAATCTTT	6180
ACAAATTGCA ACCGAAACCT TTCCCTCTGT ACCAAAGGGC TGAATAGGAT GCGGAACGTC	6240
AATTTTTTCA GAAAAACAA CTGAACTTAA TTCTTCGCGC TGTTTAATTC CATAAAGTAA	6300
AGATTGTTCA AACTCATTTG ATTCACCAAC AGATAAACTC TCAACCATCT TTTCAAGTAA	6360
ATTTACCTTG TCTGATTCAG TACATATTAA AAAGTTTTCT TTAATAAAAT ACTGTCTAAA	6420
GCCGTTGTTT TCAAAATTGT TAATCTTTGA TGATTGTACA TAACTAGAAA CTTGCATCTA	6480
ATCCATAGCT TTTCTAATCA TTTCCATCTC ATCACTCTTA AGAAACACAC TAACTTTAAA	6540
AACTGGGATT TGAAAATATA GATTTGATAA ATCAATAGCT GACACTATAA AATCTATTCC	6600
TTTAAGTTTT TCTTGATTCA ATTCATAGTA GCCTATTACA TCAACAACTT CTAATCGCTT	6660
CCCAAACCTC GTTTCCAAAC GATTTCTTAA CATTTGGGCT GCACCAATC CTGTTGCACA	6720
AATAGCAAGA ATATTAACT TAGTACTCTC TTTGCTACGT TCCATAGCAG CTAAAAAGTG	6780
AAGACTTACA TATGCTACTT CATCATCTGA TATTGTCCAC TCCAAGAACT TGTCCATATT	6840
TGCAAGAAAT TCTCTAGTCA TAAAGAATAT ATCACTATAA TTCTGTTTAA TTTTCATCTAC	6900
CAAGGGTTA TTTAAGGTAA TCCGGCTTTC TAAACGTACT TGTAGTGTCA TTAGATGAGT	6960
TATCAATCCT TCAATTAGTT GGAAATCTGA AGAAAAGTTA TACATATCAT CTAATCCTAA	7020
ATTCTGAAAT GTTTTAAATA AAGATTTTTT TAAACTTCT TCAGAAATAT TCTTCTGATT	7080
TTTTTGACAT TGTGACTCT TAGCTAACAA ATGCAAAGTA ATGTAGTCTA TTTCTGAAAC	7140
TGGAATTC TGAATTTGTTA CTTCTCTTAC TTTAGAAAGA ATTCTTTGGG CAACCTTTCT	7200
CTCTATTGCA TCATCAGTCA TCTGACAGTC TATATTTTTT ATTTCAAATC CGGATTTTAA	7260
ACGAATCACA GACAATGCTA TGTGAACTAC TAAATTCTGT AGTACAAAAT CAGATAGTTT	7320
TAGGTTGGCC TCTTGGCATT CATCCAAAAC AATTCTAGCA AATTCTTCTA ATGGAACAGT	7380
TTGATCAAAA AAGTTAAATT TTACATAGCA ATGTATTGTT TTAATAAATT GATTCTCTAG	7440
GAAATAATTT ATGATAAAAC GTCGTTTATC ACGTTCCTCG CCTGAGACAT AAACCTCTTT	7500
ATTCGCCCTA CTCTCAATGG ACAAATTATA CTCTGATAAC ATCACTCGTA TCTTTCTGAA	7560

836

ATCATGAGAT AATGTTGAAC GACTAACGTA AAGTTCATCA GCTAAATCAT CAAAAAGAAC	7620
TGGAACCTTGC TCAAATAATA ATTTATTTAA GATAAATACT AAACGATCAT CACCTTTTGA	7680
AACCGCAGTT TTCGTATAGT CTTCTTCCAG TTCATAAGTT TGTCTAAACT CCTGGTAAGC	7740
GCCTTGATTG TCAAAAAATA TTTGATACCC TTGACCTTGT TTTGAAATCA ACCGGACTCC	7800
TTGAATAATC ATTGTCTTCT CAATTAATTT CAGTACATTA CGGACAGTTC TATCTGAACA	7860
GGATAAATAT TCTGCCAGTT CTTTGCTTGT AACAAAACGT TCCTTATTTT TTATTAAAAA	7920
TTGAAGGATA TCTTTCTCTT TAATGTTTAA CACATTCATT CCTCCTAAA ACGTATGTTT	7980
TCATATATTG AAGCATATTA TACACTTAAA TCAGTTTATA TCAAACTCAA AACAATTTAT	8040
CTTAACCTAA ATATTTATTG ACATTTTCATG TGTTCATCAA ATATTCTCAA GAATCAAAT	8100
AGCCATTTTT TCAATTCCTA TTGGAATAGG AATATAGGCT TGAGGAGGTA TTTGTACAAC	8160
TGGTTTTCTT GCTTTAGAAC CAGCCTCTTC AAATTGCTTA AAGTACATTT TTGTTTGAGG	8220
ACTGACAAGA TACAAATCAA AAGCTGCTGC TCGATAGCT TTCCCTCCTT CAGTAGCACT	8280
AATAGCATCA ACTACAATAT CTTTCCCTTT TCCTTTTAGA AACTCTGTTG TTTCTGTGC	8340
CATAAGTGAT GAAGACATTC CTGCTGCACA AATAATTAAA GCTTTTGCCA TAATATTTTC	8400
TCCTTTTCTT AAATCCAATC AAAGCTGTGC TAAGTTGGCT TATTTGTTAT CTATTTTAT	8460
TATAAATAA AGCGTTTCCA ATGACAATTC CCTCATTTTC CTAAATGATA TGGAAAAAAA	8520
TTATTTATAC TTCAATTTAT AAAATAAAAT TATTCCTGAG AGTAGAAATG AAACACTATT	8580
TGCTAAAATC AAAGGCAAGT CTCCTATACG AATACCATGA GCAAGCCACA ATGCAATACC	8640
AATAACTTGC ATAACATACA TACCTAGAGC AATAGATCCT GTGTCCTTTG TCTTAACCTAC	8700
ACGAAAAACT TGTGGTAAAA ATGCAAATGT TGTTAAAATT GCTGCAATAC TTCCAATCAT	8760
ATGTCACCTC AATATGCTAA ACAAACTGAG AATAATCTCA GTTTGTTTAT ACTATTCTAC	8820
TGATTCACCG TTAGATGAAA TAACTTCCTT ATACCAGCCA AAAGATTTTT TCGGGGAACG	8880
ATTATAACTT CCCTTCCCAT TATCATCTTT ATCTACATAA ATAAAGCCAT AACGTTTCCG	8940
CATTTACCG GTACCAGCTG AAACCAAATC AATACATCCC CATGGAGTAT AACCCATTAA	9000
ATCAACACCA TCTTCAACTA CAGCCTTTTT CATTTACGA ATATGGGCAC CTAGATATTC	9060
AATTCTATAA TCATCATGTA CCATACCATC TGCTGCAACT TGATCTATAG CTCCAAAACC	9120
ATTTTCAACA ATAAAGAGTG GTAAGTGATA GTGGTCTGTA AACCAATTTA ACGCATAACG	9180
CAAACCTTCT GGATCAATTT GCCACTCCCA TTCAGAAGCC TTAACATAAT TATTTTTCAC	9240
TAAATCTTCT GTTTCAAGAT AATCAAAATA AGGATTATTT TCACGATGAG AGTCGATAGC	9300
AAAGGACATA TAGTAACTGA AACCAATGTA ATCTACAGTC CCACCAAGTA AATCTTCTTT	9360

837

ATCCTGGGCA GTAAAATCAA CTGAAATACC TTTTCGTTCC CAATACTTGA AAATATGCTC	9420
AGGATATTTA CCTAAAACAT GCACATCAGC AAAATAATAA CGCTTCTGCA TAGCTTTTCAT	9480
TGCCATTAAG ATATCCTTAG GATTGCAAGT AACTGGATAA ATTGGACACA TCGCAATCAT	9540
ACAACCTATT TGAAAATCTG GATTAATCTC ATGACCAATT TTTACAGCTC GTGCAGAAGC	9600
AACTAATTCG TAATGTGCTG CTTGATACAT AATTGCTTCT CTATTATCAC CTTCCCTCATA	9660
TACAATACCT GAGTTAGTAA ATGGTGCAAA ATCTTCCTGA TAATTCGCTT GATTATTGAT	9720
TTCATTGAAA GTCATCCAAT ATTTAACCTT ATCTTTGTAA CGTTTAAATA CGACTTCTGC	9780
AAAACGAGCA AAGAAATCAA TCAATTTCTT ATTTTCCAA CCACCATATT CGGTCACTAA	9840
GTGATAAGGC ATTTCAAAAT GAGATAGAGT GATGACAGGT TCAATACCAT TCTTTAAGCA	9900
TTCATCAAAA AGATTATCAT AAAACTGTAA TCCTTCTTCA TTCGGCTCTA ACTCATCACC	9960
TTTTGGAAAG ATACGTGTCC ATGCAATAGA GGTACGGAAG CACTTGAATC CCATTTTCAGC	10020
AAAAAGTGCT ATATCTTCTT TATAACGGTG ATAAAAATCT ATCGCCTCAT GATTTGGATA	10080
ATATTTACCC TCTAAAATC CCAAAGTAAT TTCACGAGCT ACTCCATGAC GACCAGCAGT	10140
CATAACATCA GCAACACTAA TTCCTTGCC ACCTTCTTGC CATCCACCTT CAAGTTGATG	10200
AGCAGCAACA GCACCACCCC ATAAAAATCC ATCTTTAAAA GTAGTCATCT TTTTCTCTCC	10260
TGACTTTGAT ACTCTTATTA TAAACCTTAA ACCAAAAGAT GAAAACGCAT TCTTTTCTCT	10320
TATTGTTAAG GAAAGAAGTA ATTTTAAATG GAAATAGAAC AATATCTTCT TGTATTCTCG	10380
TAATGATATC TTTACGATTT TCAATACTTT CAAACTACAA AAACCTCAC AATAATTCTA	10440
ATCCCTGTG TCTATAAAG ACTTATCGCT TTCTGGCATC CCAGAATCAT CTTCTATATA	10500
ACGTTCAACT TGCATCTGCA AGTGATATTT TTTTCTTAAA TCTAAGATTT TCTGCATTGT	10560
CTTTGATTGA TAATGTTTAT CTAAAGTTTC TTGATTTATC CACTGATCAA TAAGGAGAAT	10620
AGTTCCCTCT TTTTCAATTG GTAAAAATA TTCGTATTTT AAGTTACCTT TTTGATTCTT	10680
AATTTCTTTA ACAAGGCCAC TATCAAGCAT TTCTCTTGCA AACTTTATTG CACTATCTCC	10740
ATCACCTTTA TAATATACAT GAATAGTCAA TGTATCTTA TATCCTCCAA AATCATCCTT	10800
CAATTTTAAA AAAACAAGTT TAGATGAGGA TCTAACTTG TTTTATGA ACTAATTATC	10860
TAACGTTTCG CCATTACTTT CAATCACTTC TTTATACCAA TAAAATGATT TTTTCTTATA	10920
GCGATTTATA GTCAATTGAA ACAAGAGCAG GACAAAAGAG CCTCATAAAA GGTATTGCAA	10980
CTTGTAATA CCTTTTGTAG GTGCTTTTGT ATATGAGCCC ATGTTTCTC AATAGGATTG	11040
TACTCAGGTG AGTAGGGAGG AAGAGGTAAA AGTTTATACC CAAACTCTTC ACACAAGAGT	11100

838

TCTAGCTTCC CCATTCTATG GAATCTTGCA TTATCCATAA TAATAACCGA TGGTGTGGTT 11160  
 AATGTTGGTA AGAGAACTT CTGAAACCAA GCTTCAAAAA AGTCGCTCGT CATCGTCTCT 11220  
 TCGTAAGTCA TTGGAGCGAT TAACTACCA TTTGTTAGAC CTGCAACCAA AGAAATCCTC 11280  
 TGATATCTTC TTCCAGATAC TTT 11303

(2) INFORMATION FOR SEQ ID NO: 116:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1112 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

CCTTAGATTT CCACTTGCCA GAGGAATTGA TTGCCCAAAC GCCCCTTGAA AAACGTGATG 60  
 CCTCCAACT CCTCATCGTC AACCGTGAGA CAGGAGAAAT GCAAGATAAA CATTTCCACT 120  
 CTATTATTGA TATGCTGGAA CCTGGTGATG CCCTTGTCAT GAACGACACC CGAGTTCTCC 180  
 CTGCCCCGCT CTATGGTCAA AAAGTGGAGA CAGGAGGTCA TGTGGAACTT CTCCTCCTTA 240  
 AGAACACTAG TGGAGACGAG TGGGAAGTTC TGGCTAAACC TGCCAAACGC CTCAAGGTCG 300  
 GTACTCGTAT CAGCTTTGGT GATGGCCGCC TCAGCGCTGT CGTTACAGAA GAATTGACCC 360  
 ACGGGGGACG CATTGTCCGC TTTGAATACC AAGGAATTTT CCTAGAAGTC TTGGAAAGTC 420  
 TGGGAGAAAT GCCTCTGCCA CCTTATATCC ACGAAAAATT AGATGACCGT GAACGTTATC 480  
 AAACCGTCTA CGCCAAGGAA AGTGGCTCTG CTGCAGCACC GACTGCTGGT CTTCACTTCA 540  
 CCAAAGAACT GCTGGCAGAA ATCCAAGCTA AGGGTGTTCA TCTAGTCTAT CTGACTCTCC 600  
 ATGTCGGACT CGGAACCTTT AGACCTGTTT TGTGGATAA TCTGGACGAA CACGAAATGC 660  
 ACTCAGAGTT CTATCAACTT TCTGAGGAAG CTGCTGCCAC CCTTCGCTCT GTCAAAAAAA 720  
 ATGGTGGTCG TGTCATCGCT GTCGGAACCA CTTCTATCCG CACCTTGGAA ACTATTGGTT 780  
 CCAAGTTTGA TGGGCAAATC CAAGCAGATT CTGGTTGGAC CAATATCTTT ATCAAACCTG 840  
 GGTATGAGTG GAAGGTCGTG GATGCCTTCT CAACCAACTT CCACCTGCCA AAATCAACTC 900  
 TGGTCATGTT GGTCTCTGCC TTTGCAGGCC GTGAATTAGT CTTAGATGCC TACCACCATT 960  
 CCATCCAAGA ACACTACCGC TTCTTCAGTT TTGGTGACGC CATGTTTATT TATTGAGAAA 1020  
 GAATTTCTCT AAATCTTCTA ATACCAATAA ATCGCTAAGA TATTATTTCA AAGAACATCT 1080  
 ACAATTGAAA CTCTAGCTAG CTGTAGAAGA GGCCTAGTAC ATTGAAATTA AAATGCTTCC 1140  
 CCCTAGCTTC GAAAATATTG CCATAGATTG CGTTGACTCT CCAAATTGAT TCATCTATAT 1200

839

TTTATTTTCAG CTTCTTATAC TTTCTTCGCT GTTTGTAAAT CAAAATGCAA GACACATGAG	1260
TAGCACCATA TTTGTTACTC TTATCTGTCC TCTCAAGAGA CTATTATGAG TTATTTTCAGA	1320
ATCATTCACT ACTTTGACCC TGACTCTCCT TAGTCTCAAA ATCAAAGACT TATACTCTTC	1380
AAAAATCTCT TCAAACCGCG TCAACGTCAC CTTGGATTAT ATATGTGatC TGaCTTCGTC	1440
AGTTCTATCT ACAACCTCAA AGCAGTACTT TGAGCAACCT GCGACTAGTT TTCTAGTTTG	1500
CTCTTTGATT TTCATTGAGT ATTAACAAA AAGTGAACAA ATCTGAATTC TAATGTACAG	1560
AAGACTAGGC TTGTTCACTT TTTTATAGTC GCTATAAGAT GACCTTATCT ATAGCTTTTT	1620
ATATATAATT ATATATTCAG ACATACTATT ATCAATTTTG TCGCAGGGAG GAATCTGTTA	1680
ACGCACCCAT TCACCATTAT CATTGACTCT ATAGCCATCT ATACTTGAT TGACCGCTAA	1740
CTCACCCGAT GTATTTACAT AATACCATT ACCACCAACT TGGAAACCATT GATTGACTTT	1800
CATAGAACCG TTGCTGTTGA GGTAGTACCA TGAACATTA ACTTGATCCC AACCTGTTGC	1860
CATGGAACCA TCAGTATTAT AAAAATACCA CATACCATT TCTTGTTTCC AGTCTGTTGT	1920
TGGAGCAACT GCTTTAGCTG GTTCTACTGC TACATCTGTT CCTTGGTTAG ATGTAACAGA	1980
TACAGGATAC GAAGGAATAG ATGATTGCTC AGGAACAACA ACTTTTTCAG GTTCTCTCGT	2040
CCCTCTCCTT ATACGTCTTT TTACCATCTC TTTAGTAATT TGACGAGAAG TAGTTTCTTC	2100
AATTGTTCCA TCACGTTTAT CTACAGTATA GATTGTAGTA AGAGTAATTT ACCAATTTCT	2160
CCTACTTCTT CTACTTCTTG ACTTTTATCA AGAGTTGGGC CATCGAGATA TTCTGTTTCG	2220
ATTGGAATTT CTTGGACAAG AACTTGGGGC TTGGTTCTTT TTTTAACAAC TCTGTTTGA	2280
GAGTCTTTTT TTTGACTTAA AGTACTCTCA GTTACTTGTC CACTCTTCC ATCTACATTA	2340
TAAGTTATCG TTGTAACGTG TTTCCCATTC TTTCTAGAG TAATCTCTTG CTCTGTCTCT	2400
GCAGAAAGGT CATTGTCTGC TTCATATTTA GTAGCAAATG GAACAAGAAC TTCTTCAACC	2460
TTGCTTTTAG CTGGAACCTT GATAACTGTA TCCGTGGCTT CTTTCTATC AACAGTAACC	2520
TGTTCCGTAA CATAACCAGT CTCTGGATTA ACATCGTAGG TCCTTGTCGT AGTTACATAG	2580
CCATCCTCTC CATCAATTGT AACAGGATTT TCACTACGGT CTTTGTTC ATCTTTTCA	2640
TAACGAATTC GCGTACTTGA AATTTTCTTG GTTACTACCT TAGGTTTAGT CGCTACTTTT	2700
ACAATAATAT CCCCATGTG AGCGTCATCA TACTCTATTC CCTCTTCTT ATCTCTAGTA	2760
TCATCTCTGA CATATTGAAT CCCATCAGCA GCATGAACAA AACTTGATTT CAGATTCTCTC	2820
CTAAAAATAA AGTTAGCCCG ATTACCGCAG AACCAAAAAT CTTTCCGAGT TTACGTATTG	2880
CATAGCGCTT ATTAGTATTA GATTTTGCCA TTACATCCTA CTCTAGTAT AGCATCTTTT	2940

840

CTATCAAACG TTAAACAATA TACGTTATAT ATAAAATAGA CTTAGAATGA TATATTGATT	3000
ATTGAACTAA CACTTTAACT ATATCGTAAT CAATCTCATA TATAAAGGAT TGCAGACATC	3060
TTATCTAAAT ACATGCGAAT ATATTTAGAT ACAAACATTC CAACTTGATA AT	3112

## (2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4327 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

CCCCAAAATC TCTTCAAACC ACGTCAGCTT CGCCTTGCCG TAGTATGGTT ACTGACTTCC	60
TCAGTTCTAT CCACAACCTC AAAACAGTGT TTTGAGCATC ATGCgGCTAG CTTCTTAGTT	120
TGCTCTTTGA TTTTCATTGA GTATAAAAAC AGATGAGTTT CTGTTTTCTT TTTATGGACT	180
ATAAATGTTT AGCTGAAACT ACTTTCAAGG ACATTATTAT ATAAAAGAAT TTTTGAAGC	240
TAAATCTAC TATATTACAC TATATTGAAA GCGTTTTAAA AATGAGGTAT AATAAATTTA	300
CTAACGCTTA TAAAAAGTGA TAGAATCTAT TTTTATGTAT ATTTAAAGAT AGATTGCTGT	360
AAAAATAGTA GTAGCTATGC GAAATAACAG ATAGAGAGAA GGGATTGAAG CTTAGAAAAG	420
GGGAATAATA TGATATTTAA GGCATTCAAG ACAAAAAAGC AGAGAAAAAG ACAAGTTGAA	480
CTACTTTTGA CAGTTTTTTT CGACAGTTTT CTGATTGATT TATTTCTTCA CTTATTTGGG	540
ATTGTCCCTT TTAAGCTGGA TAAGATTCTG ATTGTGAGCT TGATTATATT TCCCATTATT	600
TCTACAAGTA TTTATGCTTA TGAAAAGCTA TTTGAAAAG TGTTGATAA GGATTGAGCA	660
GGAAGTATGG TGTAAATAGC ATAGGCTGAT GTCCATCATT TGCTTATAAA GAGATATTTT	720
AGTTTAATTG CAGCGGTGTC CTGGTAGATA AACTAGATTG GCAGGAGTCT GATTGGAGAA	780
AGGAGAGGGG AAAATTGGCA CCAATTGAG ATAGTTTGTT TAGTTCATTT TTGTCATTTA	840
AATGAACTGT AGTAAAAGAA AGTTAATAAA AGACAAACTA AGTGCATTTT CTGGAGTAAA	900
TGTCTTATTT CAGAAATCGG GATATAGATA TAGAGAGGAT CAGTATGAAT CGGAGTGTTT	960
AAGAACGTAA GTGTCGTTAT AGCATTAGGA AACTATCGGT AGGAGCGGTT TCTATGATTG	1020
TAGGAGCAGT GGTATTTGGA ACGTCTCCTG TTTTAGCTCA AGAAGGGGCA AGTGAGCAAC	1080
CTCTGGCAAA TGAAACTCAA CTTTCGGGGG AGAGCTCAAC CCTAACTGAT ACAGAAAAGA	1140
GCCAGCCTTC TTCAGAGACT GAACTTTCTG GCAATAAGCA AGAACAAGAA AGGAAAGATA	1200
AGCAAGAAGA AAAAATTCCA AGAGATTACT ATGCACGAGA TTTGGAAAAT GTCGAAACAG	1260



841

TGATAGAAAA AGAAGATGTT GAAACCAATG CTTCAAATGG TCAGAGAGTT GATTTATCAA	1320
GTGAACTAGA TAAACTAAAG AAACCTGAAA ACGCAACAGT TCACATGGAG TTAAAGCCAG	1380
ATGCCAAGGC CCCAGCATTG TATAATCTCT TTTCTGTGTC AAGTGCTACT AAAAAAGATG	1440
AGTACTTCAC TATGGCAGTT TACAATAATA CTGCTACTCT AGAGGGGCGT GGTTCCGATG	1500
GGAAACAGTT TTACAATAAT TACAACGATG CACCCTTAAA AGTTAAACCA GGTCAGTGGA	1560
ATTCTGTGAC TTTACAGTT GAAAAACCGA CAGCAGAACT ACCTAAAGGC CGAGTGCGCC	1620
TCTACGTAAA CGGGGTATTA TCTCGAACA GTCTGAGATC TGGCAATTTC ATTAAGATA	1680
TGCCAGATGT AACGCATGTG CAAATCGGAG CAACCAAGCG TGCCAACAAT ACGGTTTGGG	1740
GGTCAAATCT ACAGATTCCG AATCTCACTG TGTATAATCG TGCTTTAACA CCAGAAGAGG	1800
TACAAAAACG TAGTCAACTT TTTAAACGCT CAGATTTAGA AAAAAACTA CCTGAAGGAG	1860
CGGCTTTAAC AGAGAAAACG GACATATTCG AAAGCGGGCG TAACGGTAAC CCAAATAAAG	1920
ATGGAATCAA GAGTTATCGT ATTCCAGCAC TTCTCAAGAC AGATAAAGGA ACTTTGATCG	1980
CAGGTGCAGA TGAACGCCGT CTCCATTGCA GTGACTGGGG TGATATCGGT ATGGTCATCA	2040
GACGTAGTGA AGATAATGGT AAAACTTGGG GTGACCGAGT AACCATTACC AACTTACGTG	2100
ACAATCCAAA AGCTTCTGAC CCATCGATCG GTTCACCAGT GAATATCGAT ATGGTGTGG	2160
TTCAAGATCC TGAAACCAA CGAATCTTTT CTATCTATGA CATGTTCCCA GAAGGGAAGG	2220
GAATCTTTGG AATGTCTTCA CAAAAAGAAG AAGCTACAA AAAAATCGAT GAAAAACCT	2280
ATCAAATCCT CTACCGTGAA GGAGAAAAGG GAGCTTATAC CATTCGAGAA AATGGTACTG	2340
TCTATACACC AGATGGTAAG GCGACAGACT ATCGCGTTGT TGTAGATCCT GTTAAACCAG	2400
CCTATAGCGA CAAGGGTGAT CTATACAAGG GTGACCAATT ACTAGGAAAT ATCTACTTCA	2460
CAACAAACAA AACTTCTCCA TTTAGAATTG CCAAGGATAG CTATCTATGG ATGTCCTACA	2520
GTGATGACGA CGGGAAGACA TGGTCAGCTC CTCAAGATAT TACTCCGATG GTCAAAGCCG	2580
ATTGGATGAA ATTCTTGGGT GTAGGTCCTG GAACAGGAAT TGTACTTCGG AATGGGCCTC	2640
ACAAGGGACG GATTTTGATA CCGGTTTATA CGACTAATAA TGTATCTCAC TTAGATGGCT	2700
CGCAATCTTC TCGTGTCATC TATTAGATG ATCATGGAAA AACTTGGCAT GCTGGAGAAG	2760
CGGTCAACGA TAACCGTCAG GTAGACGGTC AAAAGATCCA CTCTTCTACG ATGAACAATA	2820
GACGTGCGCA AAATACAGAA TCAACGGTGG TACAACATAA CAATGGAGAT GTTAAACTCT	2880
TTATGCGTGG TTTGACTGGA GATCTTCAGG TTGCTACAAG TAAAGACGGA GGAGTGACTT	2940
GGGAGAAGGA TATCAAACGT TATCCACAGG TTAAAGATGT CTATGTTCAA ATGTCTGCTA	3000

842

TCCATACGAT GCACGAAGGA AAAGAATACA TCATCCTCAG TAATGCAGGT GGACCGAAAC 3060  
 GTGAAAATGG GATGGTCCAC TTGGCACGTG TCGAAGAAAA TGGTGAGTTG ACTTGGCTCA 3120  
 AACACAATCC AATTCAAAAA GGAGAGTTTG CCTATAATTC GCTCCAAGAA TTAGGAAATG 3180  
 GGGAGTATGG CATCTTGAT GAACATACTG AAAAAGGACA AAATGCCTAT ACCCTATCAT 3240  
 TTAGAAAATT TAATTGGGAA TTTTGTAGCA AAAATCTGAT TTCTCCTACC GAAGCGAACT 3300  
 AGAGAGATGG GCAAAGGAGA GATGGGCAAA GGAGTTATTG GCTTGGAGTT CGACTCAGAA 3360  
 GTATTGGTCA ACAAGGCTCC AACCCTTCAA TTGGCAAATG GTAAACAGC GACTTTCCTA 3420  
 ACCCAGTATG ATAGCAAGAC CTTGTGTTT GCAGTAGATA AGGAAGATAT CGGACAGGAA 3480  
 ATTATTGGTA TAGCTAAAGG AAGCATCGAA AGTATGCATA ATCTTCCTGT AAATCTAGCA 3540  
 GGTGCCAGAG TTCCTGGCGG AGTAAATGGT AGCAAAGCAG CGGTGCATGA AGTTCCAGAA 3600  
 TTTACAGGGG GAGTTAATGG TACAGAGCCA GCTGTTTCATG AAATCGCAGA GTATAAGGGA 3660  
 TCTGATTGCG TTGTAACCTT TACTACAAAA AAAGATTATA CTTACAAAGC TCCTCTTGCT 3720  
 CAGCAGGCAC TTCCTGAAAC AGGAAACAAG GAGAGTGACC TCCTAGCTTC ACTAGGACTA 3780  
 ACAGCTTTCT TCCTTGGTCT GTTTACGCTA GGGAAAAAGA GAGAACAATA AGAGAAGAAT 3840  
 TCTAAACATT TGATTTTGTA AAAATGGCTC TTTGTCAACT GTAGTGGGT GAAGTCAGCT 3900  
 AAGCTCGAGA AAGGACAAAT TTTGTCTTT CTTTTTTGAT ATTCAGAGCG ATAAAAATCC 3960  
 GTTTTTTGAA GTTTTCAAAG TTCCGAAAAC CAAAGGCATT GCGCTTGATA AGTTTGATGA 4020  
 GATTATTGGT CGCTTCCAAT TTGGCGTTAG AATAGTGTAG TTGAAGGGCG TTGACGATTT 4080  
 TCTCTTTGTC CTTTAGAAAG GTTTTAAAGA CAGTCTGAAA AAGAGGATGA ACCTGCTTTA 4140  
 GATTGTCTC AATGAGTCCG AAAAATTTCT CCGGTTCTT ATTCTGAAAG TGAACAGCA 4200  
 AGAGTTGATA GAGCTGATAG TGATGTTTCA AGTCTGTGA ATAGCTCAAA AGCTTGTTTA 4260  
 AAATCTCTTT ATTGGTTAAA TGCATACGAA AAGTAGGGCG ATAAAAATGT TTATCGCTGA 4320  
 GTTTACC 4327

(2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3521 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

CTCTGGCCCT GCCACTCCAA CGTTTTGTCA GGGTGCTTTT TTCATAAAGG AGTTCTTATG 60

843

TTAGATATCA AACGTATTCTG TACAGATTTT GAAGCTGTCTG CAGAAAAAATT AGCTACACGT	120
GGTGTAGATG CTGCTGTCTT GAATGAAATG AAAGAAATCG ATGCTAAACG TCGTAACATC	180
TTGGTCAAGG TTGAAACTCT CAAAGCAGAA CGTAACACAG TTTCTGCTGA GATTGCCCAA	240
GCTAAGCGCA ACAAGGAAAA TACAGATGAC AAGATTGCTG CCATGCAAAA TCTATCTGCT	300
GAGGTAAAG CCTTGGATGC TGAATTGGCA GAAATCGATG CTAAATTGAC AGAATTTACA	360
ACGACTCTTC CAAATATCCC AGCTGACAGC GTTCCTGTTG GGGCTGACGA AGACGACAAT	420
GTGGAAGTTC GCCGTTGGGG TACTCCACGC GAGTTTGACT TCGAACCTAA AGCTCACTGG	480
GATCTCGGTG AAGACCTTGG TATCCTTGAC TGGGAACGCG GTGGTAAGGT AACAGGCGCT	540
CGCTTCCTCT TCTATAAAGG CCTCGGTGCT CGTTTGAAC GTGCTATCTA CAACCTTTATG	600
TTGGATGAAC ATGGAAGAAG AGGCTATACT GAAGTCATCA CACCTTACAT AGTCAACCAT	660
GATTCTATGT TTGGTACTGG TCAGTATCCA AAATTTAAGG AAGATACTTT TGAATCAGC	720
GATACCAACT TTGTCTTGAT TCCAACCTGCT GAAGTTCCTC TGACAACTA CTACCGTGAT	780
GAAATCTTAG ACGGCAAAGA TCTTCCAATC TACTTCACTG CCATGAGTCC GTCATTCCGT	840
TCTGAGGCTG GTTCTGCCGG TCGTGATACG CGTGGCTTGA TCCGTTTGCA CCAATTCCAC	900
AAGGTTGAAA TGGTCAAATT TGCCAAACCA GAAGAATCTT ACGAAGAATT GGAAAAAATG	960
ACAGCCAACG CTGAAAACAT TCTTCAAAAA CTCAACCTTC CATACCGTGT CGTTGCTCTC	1020
TCTACTGGAG ATATGGGCTT CTCAGCTGCG AAGACTTACG ACTTGGAAGT GTGGATTCCA	1080
GCACAAAACA ATTACCGTGA AATCTCAAGC TGTTCAAACA CAGAAGATTT CCAAGCCCGT	1140
CGTGCCCAAA TCCGTTACCG TGATGAAGCA GATGGCAAGG TGAAACTCCT TCATACCTTG	1200
AACGGTTCTG GACTTGCAGT TGGACGTACA GTGGCTGCAA TTCTTGAAAA TTACCAAAT	1260
GAAGATGGTT CTGTGACCAT CCCAGAAGCA CTTCGTCCAT ACATGGGTGG AGCTGAAATC	1320
ATCAAACCAT AAAAAATAAG GTTTAGCTAT TTCTAGCTAG ACCTTTTTTC GTAACCAAAT	1380
CAGATAAGCA CCTAGTACAA AGAATAAAAT AGTTAGGCAT ATAATGGTTT CAGCCAATAC	1440
CAGGTAATCC AGAAATGGAA GTTTCAAAAT TCCTGAGCC ATCTTGAGCG AGGTCGCTGT	1500
GATAATGGTT GGAAGGTGA GGGCTGAGAA GGCTGGTTGA AAACCTTGTT TTAATATGTT	1560
GGGCAGACGA GTTAAACAA AGAAAAAGAA GGATTGAGAA GCCAAATCA TGACAATCAA	1620
GACCCAAGTC GGCAGGCTGG TTCCTCCTAC TCGAACTAGA GAAGCCAAGA GTAGAGAGAA	1680
AGGAGCACAG TAGATTCTTT CTGTCCAAG CAAGGCTAGT GGGAGTGGAT GTTCTTTTAA	1740
ATCGCTATAA ATAAGGGGAT AGAGATAGAA GGTCAAGAGA AAACCAAAAC TCAAGGTCGC	1800

844

ATAGGCAATT TCGATAATAC CTACCAGAGG ATAGGTCAAG GCAGCCACTG CTATCCCCAC	1860
ATAGAGAACC GTCCAGCTTG GAGTGGCATG AACCCCTCCGC CCTGGACAAG CAAACTTGAT	1920
GGTAAACCA GCAATCAAGG TCAAATCCAA GAGAAATGAA AACCACCAA TCCCTTGTGC	1980
TACCAAAGGA AGATAAGAGA ATACGCGAAA GACATAGGTC GATAAAATCA TCCAGCCAT	2040
AGGAAAGGTT GCCATTCTG ACAAAGAGG GGGCTTGGTC AATTCTTGCT TGGTTTCTTT	2100
CCAATTAAAG AGATGCAGAA TTAGAAAGTA AATCCATAAA ACCAAACCAA TCAGACTAAA	2160
AAGATGGGAT AGAACCGGCA ACGTATCTAA AATAAGATTT CCAGCTCCTG CCAAACCTAG	2220
CAAACAACCT GAAAATACTA AGGGGAGTTT TTTTCATCCTA ACCTCCAATA ATCATGTTAG	2280
TTTCAGTATA ACATAAAAGC GCTTAAATGA GGATTTAAAA AAACGAGTCC GCTTATTTCA	2340
GACTTCATTT TACTCAGATA TGAATTAGGC ATAAGGTTGC AATTCTGGAT TAATTGGTGT	2400
ATTAGCTAAG TTGTTGGCAT ACTTACAGAG GATTGCTAGG CTGACACCAA AAACCACATC	2460
CAAGGCATTT TGTGAGTGT AGCCAGCTTC TAAAAACTCA GACAAGGCTT CATCTCCTAC	2520
ACGACCCTTG GTATTGATAA CTGCCAAGGT AAAGTTAGCT AGGGTATCCA ATTTAGGATC	2580
TGTTTCAATT GGAGTACGAT TCGGAAGAGC TTGAATCAAG TCATCATTCA TCTGGATTG	2640
TTTGATGGAA AAGGCTGTGT GACCTGCGAC ACAGAAGGCA CAACCATTGG TCACGGCTGC	2700
CGTGATTGTC ACCACTTCAC GCTCAACGGG TGTCAGGCTG TTGCGACGGT GGATAGATGA	2760
GACAATTTGG TAGGCTTCTA AAACAGTCGG GGCATTGGCC AAGAGACCGA TTAGGTTGGG	2820
AATATAGCCA TTGTTGTCTT TTTCTACTGT TTCAAGAATT TCTTTCACTT CTGCTGGTGC	2880
TGACTCTACT GTATGGATAG TAAATGTTGT CATAAGATAC CTCTTTTCTT ATTATTGACA	2940
CTAATATTAT TGGAAAATCT TATAAAATCC TGATTCTTAA GTTATCTAA GATAAAGCTT	3000
TATTCTCTCA TAAGATTTTC GTTGTATAT TAGTTTATCA CACTTCCAAT CACTTGTATA	3060
ATATATATTA TATATCAGGC TGATAAAAAT TATTTATAGG CAAAAAATC ACACGAGCTG	3120
TGTGATTCCA TTATTTGTCA AAATACTTTT TAGTTTCAGC AATAACGACT GCGGACAAGA	3180
CCAAGAGGGC AATCAAGTTT GGCAGAGCCA TCAAGGCGTT AACGATATCT GCGATAATCC	3240
AGACCATATC CAACTCGATA AATCCTCCTA ACAAGACCAT GAGCACAAAA ACCACACGGT	3300
AGAGCCAGAT AAAGCGAACC CCAAAGAGGA ACTCAAAACA GCGTTCTCCG TAATAGTTCC	3360
AACCTAGAAT CGTTGTAAAG GCAAAAAGTA CAAGGAAGAT GGTCAAGAGA GCAGGCCCAA	3420
ACTGTGAAAA GTTTGTTGAG AAAGCTGACT GAGTCAAGGC AACCCCATTC AAGTCACCGC	3480
TCCAACTCC AGTTACCAAG ATGGTCAAAC CAGTTAGAGT A	3521

(2) INFORMATION FOR SEQ ID NO: 119:

845

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1968 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

AACCTGGGCA AGCAAGCTAA AAGCAATGGG ACCTGGAATC CTAATGGCAA CTGCCGCTGT	60
TGGAGGTTCC CACATTGTAT CCTCAACTCA AGCTGGCGGT TCTTACGGTT GGTCTCTACT	120
TCTCTTGGTC ATCTTAGCCA ATGTCTTTAA ATATCCATTT TTCCGTTTGT GTGCTGAATA	180
CACAGCTGAT ACTGGAAAGA CTTTGGTTGA AGGTTATGCC GAAAAAGGAA AACTCTATCT	240
CTGGATTTTC TTTATCCTCA ATGTCTTTTC GGCTATGGTC AACACGGCTG GTGTTGCCAT	300
TCTGTGCTCA GCTATCATCG CCAGTGCCTT CCCAATGATT GGACTTAGCA TTA CT CAGTG	360
GTCCCTCATT CTCGTTGCAA TCATTTGGGC TATGCTACTC TTTGGAGGCT ACAAAC TTTT	420
AGACGGCATG GTCAAATGGA TTATGTCTGC CTTAACCATT GCGACTGTTT TGCAGTTAT	480
CATTGCGGCG GTCAAGCATC CAGAATACAG TTCTGATTTT GTCGAGAAGA CACCTTGCCA	540
AATGGCAGCT CTGCCCTTCA TCGTCTCCCT CCTAGGATGG ATGCCGGCTC CTATTGAAAT	600
TT CAGCCATC AATTCAC TTT GGTCAGCTGA AAAGAGAAAG ACCGTCAACT TTAACACAGA	660
AGACGCTCTG TTTGACTTTA AACTGTTTAT TATTGGAACA GCTATCCTAG CCGTCTTCTT	720
TGTGGCACTG GGAGCACTGA TTCAGTATCC TACAGGGCAG GCGGTTGAAG CTGCTTCAGC	780
CAAATACATC TCTCAATTCG TGGGCATGTA TGCTCTGTT CTTGGCGAAT GGTCCCGTTA	840
CTTGATTACC TTTATTGCCT TCCTCTGTAT CTTTGAACA GTTATAACTG TTATCGATGG	900
CTATTCTCGC GTTAATCAGG AATCTCTCCG ACTGCTAATC AGTCAAAAAG AGGACAATCG	960
TAAATCTTTG AACATCTGGA TGACCATCAC TGCTATCATC GGTATCGTCA TTATCAAGTT	1020
CTTCGCTGGT CAGGTTTCAA CCATGCTCCG CTTTGCCATG ATTGGCTCTT TCCTGACAAC	1080
ACCTTTCTTT GCTCTTTTGA ATTACGCCTT GGTAACGCGT GAAAACAAA ATCTTCCTTC	1140
TTGGCTCAAA CACCTTGCCA TTGCGGGATT GATTTTCCTC TTTGCTTCGC CATCTTCTTT	1200
ATCTACGCAC TCGCAATCGG AAAAGCAGGG TAAGGGACAA GCGCGAGATG AAGATAAGGT	1260
TT CATTTC AA GAGAAAATTC AGCAAATATT TCTATGATAA AAAGCATAAG AACAAGGTTT	1320
TGAAGACCTG AACTTATGCT TTTTACGTT CTAAAGACT GTTTATACTC AAAAAACAGT	1380
TGAACAACTT CAACCACCTC TTATAAGAAC TTTATACTAT TCGAGAATCT CTCAAACCA	1440

846

CGTCAGCTCT ATCTGCAACC TCAAAGCTGT GCTTTGAGCA ACCTGCGACT AGCTTCCTAG	1500
TTTGCTCTTT GATTTTCATT GAGTATTAAT TCTCCTTTC CAACTCATAC AAATCTGCGA	1560
TAATAGCTGC GACATGTTTG ATATCTTCCA GCATGCCTCG CATTTCAAAG TCAGCCAATA	1620
CAGGGAAGCC AAAGCGTTGA CTGTATTGCT TGGCTGTTAG GCAGTATTGG TTATTAAAGT	1680
TACGATTTCC TGACCCAACC ACACCAAAAC ACTTACTAGC ATTGTTACCA TAGGCAATAA	1740
AATCTCCCAC CGGTGTCGTC AAAATCTCAA CATCTCCGTT ATCCACGCCA TTCCCACCTT	1800
CGAGATAGGT CGGCAAAAAA GCGACATAGG GATGGTCCAT TTCATAGAAA TTTTTCCTT	1860
CCTTGACCAA ATCCTTGATA TGAATCTTTT GAACCTCAAT CCCTTTGTAC TGGGACAAGA	1920
GATAGTCTTT CAAGCGCGTC ACAAACCTTT CAGTGTGCCC ACTCAAGG	1968

(2) INFORMATION FOR SEQ ID NO: 120:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7172 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

CCGCATTTTT TATCACTAGA CTCGAGACAT CTTTGTAGTG GCTCTTGCTC TCTGGTTTAA	60
TTTTCTTCCT TGCTCAAGGA CTCCTGCTAT TTCTCTGGT CGTCCGACTC AAACATCAAT	120
TGCTGAGAT TTATCCTCAA ATCAATAAAA AGATTCGCTT CTAATATTTA GGGGTTCTCA	180
CCATTGATTT TCTATTTTTT GTTCTCTTAG CTTTCATTAG TTCTCAGCGT TTTTCATCTC	240
TTATGCCAAT CATCACTGCT TGCCATTCTA CTTTTATTA TATGACAGCT GACTACCTAA	300
GAGAAACTA TCCAGACTTT TACGACAAAC ACATCTCTTT ATGGGAGTGT CTCTAAAGAA	360
AAGGAGGTTT TAGCATGAAA AAAATCATCT TCATCAAAAC CATTCAACTC CTTGTCAATG	420
ATGGAATCAT GCTGGCATTT TTGACATTTA AAAGGGGGCT TACTTGGGAC TGGATTTTGA	480
TTTATAGCGG TTGGCTCATT TTCTTTTCATC CTGTGCTATT GACCTATCTT TCAAACCAAC	540
TTTGTGACCA CTTTAGTTAA CTCTATTCCC AGATTAGACC GAGATTCTGG CGTTTTGCTT	600
TACAAATTCT CCTATGGGAT AGCCTGATGA TTCTCTCCTT GGTGTCTTTA AGTGATATTC	660
CACTTTTCTT TCAGGGAAC CTCTCATCC TAGGACATCT CATCCCTTCC TATCGCATCT	720
GCCAAAGCCT GAAAAGAGAC TTCCCCCAAG CATATCAAGA ACCGATTTCT TTTTGGAGTA	780
TTTTATGATA GATGAGAAAG ACCAAGCCGA CTGGGCTTGG TCTTTCTTAT CTCTTTTATG	840
TATCTAGGAT AATGGTAACA GGTCCATTAT TAACCAGCTC AACCTGCATA TCTGCTCCAA	900

847

AGATGCCTGT CTGAACGGGC ACTTCTTGCG CTAATTTTGT ATTGAAAGCA TCATAGAAGT	960
CTGATGCCAT ATCAGGTTTA GCTGCCCCTG TAAAGCTGG ACGATTGCCT CTCTTAGTAT	1020
CCGCAAAGAG GGTAAACTGA GAAATAGAGA GGATTTCTCC TTCAATATCT TTGACAGACA	1080
GTTTCATCTT GCCTTCTGCG TCTGAAAAAA TCCGCATATT GACCAGTTT CTCACAGCAT	1140
AGTCCAAATC TTCTCTTGG TCCTCTGGTC CAACACCAAC CAGCAATAAA AGTCCCTGAT	1200
TGATTTTTCCTGCTGAATCTGG CCTTCTATAC TCACTTGGGC TTTTTTAAACC CGTTGGATAA	1260
TGATTTTCAT AATAGCCTTT CTAGTAAGAG CTAGGACAAC TAGCCGTTGG TCCGTTTGAC	1320
AGAGTAAACT TCTGGCACAC TCTTAATTTT ATCGACAACC GTGGTCAGTG TAGAGAGGTT	1380
GGCAATACCG AAGGACACAT GGATATTAGC AAAC TTCATA TCCTTGGTTG GTTGGGCATT	1440
GACCGTTGAA ATATTCTTGG TTGTATTTGA AAGA ACTTG CAGTACATCGT TCAACAGTCC	1500
TGTACGGTTG AGACCGTAGA TATCGATATG GGCCATATAC TCCTTATTTG AGCTAGGGTA	1560
CTGGTCTTCC CATTCCACAT CAAGGAGACG TTGCTCGTAG TTTTCTTGGG CACGCAGGTT	1620
CATACAGTCC ACACGGTGAA TAGCCACACC ACGACCCTTG GTAATGTAGC CAACAATATC	1680
GTCACCAGGC ACGGGGTTAC AACACTTAGC AATCCGCACT AGGAGACCAG AAGCACCTTC	1740
AATAACCACT CCCCCCTCAT GCTTGACCTT GAGGGTTTCT TTATTTTCAA CCTTGACCTC	1800
GCCACCTTTG ACAAGCTCCT CTGCCTCAGC TTTGGCCTTG GCACGCTCTT CCTCACGGCG	1860
TTCTTTTCA GTCAGACGGT TAAAGACGGT AATCGCACCG ATTTCCCCAA AACCAATGGC	1920
CGCAAAGAGG GAGTCTTCTG TCTTGTAAGT GGTCTTTTGC AGAACTTGAT CCATGTGGCG	1980
CTTGTCATA AATTTATTTG CCACATAGCC ATTTTCTTGG AACTGAGCCA TCAGCATCTC	2040
ACGACCCTTG TTGACAGACA ATTCCTTATC TTGGTTTTTA AAGA ACTGGC GAATCTTATT	2100
GCGCGCCTTG CTAGTCTTGA CCATATTGAG CCAGTCACGG CTAGGTCCAA AGGAGTTCGG	2160
GTTGGCGATA ATTTCAACCT GATCCCCTGT CTTTAACTTG GTTGTCACTG GAACCATGCG	2220
GCCATTGACC TTGGCACCAG TTGCTTTTTC ACCGACCTTG GTATGGATT CTAGGCAAAA	2280
ATCAATCGGT CCTGAATCTT TGGGAAGGGA ACGGACAGCT CCATCTGGGG TAAAAACGTA	2340
AATCTCCTCA GCCAAATAGT TTTCTTAAC AGAGTCCACA AATTCCTTAG CATCATCAGC	2400
CTGGTCTTGG AGCTCCATCA TCTCCTTGAT CCAGTTTATT CCAATAGCTG ATTCCTTGCT	2460
GTAACTTGC CCTTTTATAC CTTTCTTATA AGCCCAGTGA GCCGCAACCC CGTACTCAGC	2520
CACCTCGTGC ATTTCTTGG TTGCAATCTG GAATTCATC GGCCCTTTTG GTCCATAAAC	2580
AGTCGTATGG ATAGACTGAT AACCATTGGC CTGCGGTTG GCGATATAGT CTTTGAAGCG	2640

848

ACCTGGCATC GGTTCCTAAA ATTTCATGCAC GTAAACCAAGC ATGGCATAAA CATCACTTTG	2700
GGTATCTAAA ATACAACGAA TAGCAATCAG ATCATAGATT TCCTCAAACC GTTTTCTCTT	2760
GTCCTGCATT TTGCGGAAAA TTGAGTAAAT ATGCTTGGGA CGACCATAAA TCTTCCCTTT	2820
CAAGTGACGT TCTGTCGTAT ACTCCTCTAA TTTGTGACT ACCTCATCCA CCAAGGCCTC	2880
ACGCTCCCTG CGCTTTTCCT TCATCATATG GGTAACTCTG TAAAACTCCG TTGGATTGAG	2940
ATAACGGAAA GACAAGTCTT CTAATTCCTA TTTGACACTG GAAATCCCCA AACGATGGGC	3000
AAGCGGGGCA TAGATTTCCA TGGTTTCTTT GGAAATACGC TCCTGCTTGT CTTTTCGAAG	3060
ATGTTTCAGG GTCCGCATAT TGTGCAAGCG GTCAGACAGT TTGACCAAAA TAACGCGGAT	3120
GTCCTCAGAC ATGGCCATGA GCATCTTGCG ATGATTTTCC GCTAATTGCT CCTCGATCGA	3180
TTTGTAATCG ACCTTGCCAA GCTTGGAAC TCCGTCAACA ATCATCCGCA CATCAGGACC	3240
AAACTCTCTT TCCAAATCGT CCAAAGTCGC ATCTGTATCT TCCACCACAT CATGCAAGAA	3300
TCCACAAGCT ACTGTTACAG CATCCAGCTT TAGCTTAGCT AAAATACCTG CCACTTGGAT	3360
AGGGTGAATG ATATAAGGCT CGCCTGATTT GCGATATTGA CCACTGTGGC ATTCAACAGC	3420
ATAGACCAAG GCCTTATGGA CAAAATGAAC ATCCTCTTCC GTTAAATATT CTTTGGTTAA	3480
AGCGACAAC TCTTCGCCCTG TTAATTCAC TTCTTTGCGC ATCTCTACTC TCCAATTCTT	3540
CCTACCATT TATCACTTTT TTAAGAATAT GAAAAC TAGA TTGGAACAGA ATAAGAAAAA	3600
AATAATTCAA AATTGCTTGA TAATTCTGAA TTATTGGTCC GTAATATACT ACGAAGTTAG	3660
ATTTTAAACT TAGGTGATAG AAGGAGAGAT AGAAGAACGG AAACCATATT GTAACCCAAA	3720
GACTTTCTGA CTTCCCAAT TCCATTGAAG ATACGAAAGA TAAACGGTGG AACTCGTATC	3780
ACATACACTG GTACCTTGAC TGGATTTTGG AATTAATACT AAATGAAAAT CAAAGAGCAA	3840
ACTAGGAAAC TAGCCGCAGG TTAATCAAAG CACCGCTTTG AGGTTGCAGA TAAAGTTGAC	3900
GCGGTTTGAA GAGATTTTGG AAGAGTATAA AAATCCTCAA GATACTTTCT TCTATCCTTT	3960
AGTTTATAAG GAGAATACCT ATGAAAAAAA CTGCTATTTT TATCTTTGCT CTCCTAATGT	4020
TAGGAGTTTG CTGCCTGTTT CTATTCAGCC AGCAAAGCTA TAAAAACAG TCGTTCAATA	4080
CTATGCTAAC GACCAGAACC TGCCCACTAG GATACTTAT AGTGAATATA GCGACAAATG	4140
AGAAGCCAAC TACGGTAGCA CTCTAAACAT CACGTCTATC AAACAAGCTA ATGACGGAGT	4200
TTATGCAACC TATGAAGGGC AATTGACACC TTTCCAATAT TGATAAATTG ATAACCAGCC	4260
TGCTTTCATC TAGTCATGCT GGTTTTAAAG TTCATTTTAA ATCCTTACCT ATTCTCCCTA	4320
ACTGTGCTAT ACTTAATTTA TACTCAATGA AAATCAAAGA GCAAAC TAGA AAGCTAGCCG	4380
CAGGCTGTTC AAAGCACTGC TTTGAGGTTG CAGATAAAGT TGACGCGGTT TGAAGAGATT	4440



849

TTCGAAGAGT ATTAGTACAT TCTTTGAGAT TGGAGCTAGT ATGAAAATCC ATAAAACCGT	4500
GAATCCTGTT GCCTATGAAA ATACCTATTA TCTAGAAGGC GAAAAGCACC TCATCGTCGT	4560
CGATCCTGGT AGTCATTGGG AAGCCATTCG TCAGACAATC GAGAAGATCA ACAAACCGAT	4620
CTGTGCTATT CTCTTGACCC ACGCCCATTA TGACCATATC ATGAGTCTGG ACTTGGTTCG	4680
CGAGACGTTT GGCAATCCTC CTGTCTATAT CGCAGAGAGC GAAGCCAGCT GGCTCTACAC	4740
TCCTGTCGAT AATCTCTCCG GTCTCCCTCG CCACGATGAT ATGGCAGATG TGGTCACAAA	4800
ACCTGCAGAA CACACCTTTG TCTTTCACGA AGAATACCAA CTAGAGGAAT TTCGTTTAA	4860
GGTTCTACCG ACCCCAGGGC ACTCTATCGG TGGTGTTCCT CTAGTCTTTC CTGATGCTCA	4920
TCTAGTCTTG ACGGGAGATG CTCTATTCGG CGAAACTATC GGACGGACCG ACCTTCGGAC	4980
TGGTAGCATG GAGCAACTCC TTCATAGTAT CCAGACCCAA CTCTTCACCC TACCAAACTA	5040
CGATGTCTAT CCAGGACATG GTCCAGCTAC TACTATCGCT CACGAAAAGG CCTTCAATCC	5100
CTTTTCTAG CAAGATGATG ACAATCGAAA TTTAAGTAAA CTATCCAGCA AATCTTTCTA	5160
TTACAAAAGG CATCCTATCA AGGTTTTTAC ACATGATTGG ATGCCTTTTT TCTGATGACT	5220
AGATTTTTTG CATTACCAA TAATCACGCG CTCCTCTGGT GAACGCCACA TTCCGTCTCC	5280
TTCTTTGACA TCATAGGTTG TAAAGAAATC GTCGAAGTTT GGTACTTGCA CATTGACACG	5340
GAGTTTGGCT GGTGCGTGCA CATCGACGCT AGCCAAAAGT TTCATAAATT CTGGTCGACC	5400
TTTCATGCGC CAGATGCGAC CGAAGTTGTA GAAGAACTCT TCTGCTGAGA AGTCTGCTTC	5460
TCTCTTAGCT GCTTCAAGCG CTGCTGCGAT TCCTCCCAAG TCAGCCACGT TTTCTGATAC	5520
AGTCAATTTA CCGTTAATGG TTGCTCCATA AGAATCCTGT CCATCAAATT GGTCAATGAC	5580
TTTTTGTGTT TTCTCCTTGA AGGCAGCATA GTCGCTCTCT GTCCACCAAT CCTTGAGGCT	5640
ACCATTTTCG TCAAAGGAAG CCCCCTTAGT ATCAAAGGCG TGGGAAATTT CATGGGCAAT	5700
CACTGCCCCA ATACCACCGT AGTTAGCAGA AGATGACTGA TGCAAGTCAT AGAAAGGCGC	5760
CTGTAAAATG GCCGCTGGAA AGACAATCAG GTTCTTCTGA GGATTGTAGT AGGCATTGAC	5820
CATATGAGCA GGCATGCCCC ATTCCTTATA ATCTACAGGC TGGTTCCACT TACTCCAACT	5880
GTGCTTGATT TCCACACGCG CAAAGGCTAG AGCATTCTCA AAAAGACTGG CAGTTTCATT	5940
CACTACCTTA TCCTTGTAAC GTGCAGGCAA TTCTTCTGGA TAGCCAATAT AAGGTTTGAT	6000
CACATTGAGC TTCACGATAG CCTGTTTACA GGTTTCTGGA GTGAGCCAGT CATTCTTAAG	6060
CAGACGCTCC TTATAAACAT CAATCATGGT TGCCACTTTT TTCTCCACAT CCGCCTTGGC	6120
TTCTGGAGAG AACTTCTCAC GGGCGTACCA AAGACCCAGG GCTTGCTTGA AAGGTTCTTG	6180

850

TGCTAGATGA TAAGCTGCTT TGACCTTATC TTTTGCCTCT GGAAGTCCAG AAAGGGCACG	6240
GCTGTAGGCA CCAGACAAAA CACGGATATC CTCTGTAAAG TAGCTGGTTG AAAGATTGAC	6300
AACACTCAAA ATCAAGGTTG CTTTAAGGAG AGACCAGGCT TCCTCACTGT AGAATTGCTC	6360
TGCTGCTTGC CAGAAACGTT CCTCGTCTAC AATAACCTTG TCTGGTAATT GCCCAATAAC	6420
TGCTTTGAAG AAGTCATCCA AAGGTAGGGC AGGCGCGAAT TTCTTGAAAT CTTGTAAGA	6480
ATATGGATGA TAGAGTTTAG CATATTCTGA ACTTTCTTCA TTAGAGAGCA CCACTGCCGC	6540
AACTCGGCGG TCCAATTCAA GTCTTTTTC TAGCAAGTCT TCAATTTCTT CATCAGAGAA	6600
ATCATAAGCC TTGAGGAGAT TTGCGCTGCT TTCTTTCCAA AGAGTCAAGA GCTCTTCGCG	6660
CTGAGGATGT TCTTCTGCA? AGTAGGTCGT ATCTGGCAAG ATTGTGCTTG GAGCGCTAGC	6720
CCATAGAACA TTGATTCTAG CATCCATAAA GTCTGGCGAT ACACCAAAAG GAAGGAAGTT	6780
TGGTTTTCCCT GCAAGCTCAA ACTCTGCTAG TTTAGCTGTA AAATCCGCAA AAGTCTCCAA	6840
TTCTTGGAAT TCTTTAAGGA GTGGTAAGAC AGGTGTGATA CCGTCAGCTT CTCTCTTGTC	6900
AAAATCACGA ACTAGGCGGT GGTATTTGAC AAAGTTTTCC AAGATAGCAT CCTCAGGCAC	6960
TTCTTCACCT GCTAACCACT TGTCTGTTGT CGCCAGCATC AGGTCTTCAA TTTCCTGGTC	7020
TAAATCAACA AAACCTCCTG TTTGAGACTT ATCTGCTGGG ATTTAGCTG TCTGTTGCCA	7080
TTCTCCATTG ATAGCATCAT AAAAAATCATC TTGATAACGT GTCATCTTGT TCTCGCTTTC	7140
ATTTGTATTT GCATTTATCT TAACAAAAAT CG	7172

(2) INFORMATION FOR SEQ ID NO: 121:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4518 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

CGGGAAGTTA TGCGATCTAG ACTTCGTTCC TGTACAGCTA CTTTCTCAGG TGGTCTTGTT	60
GTTTGATGA GTTTGTTTAG AGAGGATCTT TCTATGTCTT TCTTCTTAT TTTTGTTTGA	120
TATGCTTTTC TGATTCTTA TCTAATTTAT GGTATTTC A GACTAAAAAG GAAATACCGA	180
GTAGATGAAT AGCAAGGTTT TAGGTCTTCA GATTGATTTT TAGCACTCTT GATAAAAGAG	240
TGCTAATTTT TTGAGTTTTT GTCTTGACAT TCTCTTCTAA GGGTGTATAA TAGAATCATG	300
AGTTAGCACT TGGATGCATT GAGTGCTAAT TGATCAGACA GAGAGGAGTG ATGAGATGGT	360
TACAGAGCGT CAGCAGGATA TTTTAAATCT GATTATTGAC ATCTTTACCA AAACGCACGA	420

851

ACCTGTCCGA TCAAAAGCCT TGCAAGAGTC TATTAAGTCT AGCAGTGCAA CCATTCTGTA	480
TGACATGGCG GAACTAGAAA AACAAAGGGTT GCTTGAGAAG GCTCATACTT CAAGTGGTGC	540
GATGCCAAGT GTTGCTGGTT TTCAGTACTA TGTGAAACAC TCACTGGATT TTGACCGGCT	600
GGCTGAAAAT GAGGTATATG AGATTGTCAA AGCCTTTGAT CAGGAATTCT TCAAATTGGA	660
GGATATTCTG CAAGAGGCTG CTAAGTTACT AACAGACCTG AGTGGCTGTA CGGTAGTGGC	720
ACTGGATGTT GAGCCGAGCA GGCAACGTTT GACAGCCTTT GATATCGTTG TTTTGGGGCA	780
ACATACAGCC TTGGCGGTAT TTACCCTAGA CGAGTCGCGA ACGGTTACTA GTCAGTTTCT	840
GATTCCAAGG AACTTCTTGC AGGAGGATTT GCTGAAACTG AAGAGCATCA TTCAGGAACG	900
TTTCCTCGGT CACACCGTTT TAGATATTCA CTACAAGATT CGGACGGAGA TTCCGCAGAT	960
TATCCAGCGT TACTTTACAA CAACGGATAA TGTATCGAT CTCTTTGAAC ACATCTTTAA	1020
GGAAATGTTT AACGAAAACA TTGTGATGGC GGGCAAGGTC CATCTCTTGA ATTTTGCCAA	1080
TCTAGCAGCC TATCAGTTCT TTGACCAACC GCAAAAGGTG GCCTTGGAGA TTCGTGAGGG	1140
GTTGCGTGAG GATCAGATGC AAAATGTTTC TGTTCGAGAC GGTCAAGAGT CCTGTTTAGC	1200
TGACCTAGCG GTAATCAGTA GTAAGTTCCT CATTCTTTAT CGGGGAGTTG GAATTCTAGC	1260
CATTATCGGT CCAGTTAATC TGGATTACCA ACAGCTAATC AATCAAGTCA ATGTGGTCAA	1320
CCGTGTTTTG ACCATGAAGT TGACAGATTT TTACCGCTAC CTCAGCAGTA ATCATTACGA	1380
AGTACATTAA GATTGAAATC ATTAAAGGAG GCGAACATGG CCCAAGATAT AAAAAATGAA	1440
GAAGTAGAAG AAGTTCAAGA AGAGGAAGTT GTGAAAACAG CTGAAGAAAC AACTCCTGAA	1500
AAGTCTGAGT TGGACTTGCC AAATGAACGT GCAGATGAGT TCGAAAACAA ATATCTTCGC	1560
GCTCATGCAG AAATGCAAAA TATCCAACGC CGTGCCAATG AAGAACGTCA AACTTGCAA	1620
CGTTATCGTA GCCAGGACTT GGCAAAAGCA ATCTTACCAT CTCTTGACAA CCTTGAGCGT	1680
GCACTTGCAG TTGAAGGTTT GACAGATGAT GTGAAGAAGG GCTTGGGGAT GGTGCAAGAA	1740
AGCTTGATTC ACGCTTTGAA AGAAGAAGGA ATTGAAGAAA TCGCAGCAGA TGGCGAATTT	1800
GACCATAACT ACCATATGGC CATCCAAACT CTCCCAGCAG ACGATGAACA CCCAGTAGAT	1860
ACCATCGCTC AAGTCTTTCA AAAAGGCTAC AAAGTCCATG ACCGCATCCT ACGCCAGCA	1920
ATGGTAGTGG TGTATAACTA AGATATAAAG CCCGTAAAAA GCTCGCAGTA AAAATAGGAG	1980
ATTGACGAAG TGTTGATGA ACACAAGAAA ATCTATCTTT TTTACTCAGA GCTTAGGGCG	2040
TGTTGATTC GGCAATTCTG ACGGTAGCTA AAGCAACTCG TCAGAAAACG GCAATCGCTA	2100
TGGCGTTTGC CTAGCTTCCT TACTAAGTCTG TCGTCGAAAT AAAATCGATT TCGACTCCTC	2160

852

GTGTCGCAAT TTACATAATA GAAAACCTGT CCGAAACGAC AATAAACTAT GAAGAAAGAT	2220
AAAATATGTT TGGCTTTGTA ATAGTGAGCG AAGCGAACCA AACACGATAC TCTTCGCCGT	2280
GGCGCTATTT GCGCAAATTT TGAGACCTTA GGCTCAAAGT TTAGTCAAAG AGATTGACGA	2340
AGTCAAGCTC TGACGGCGTC GCCACTGTCT CCACTTAAGA AGAGTATCAA AAAGAAAAAT	2400
AGAAAATTAA CTAACAAGGA GAAAAACACA TGTCTAAAAT TATCGGTATT GACTTAGGTA	2460
CAACAACTC AGCAGTTGCA GTTCTTGAAG GAACTGAAAG CAAAATCATC GCAAAACCAG	2520
AAGGAAACCG CACAACTCCA TCTGTAGTCT CATTCAAAA CGGAGAAATC ATCGTTGGTG	2580
ATGCTGCAAA ACGTCAAGCA GTTACAAACC CAGATACAGT TATCTCTATC AAATCTAAGA	2640
TGGGAACTTC TGAAAAAGTT TCTGCAAATG GAAAAGAATA CACTCCACAA GAAATCTCAG	2700
CTATGATCCT TCAATACTTG AAAGGCTACG CTGAAGACTA CCTTGGTGAG AAAGTAACCA	2760
AAGCTGTTAT CACAGTTCCG GCTTACTTCA ACGACGCTCA ACGTCAAGCA ACAAAGACG	2820
CTGGTAAAA TGTCTGCTTT GAAGTAGAAC GTATTGTAA CGAACCAACT GCAGCAGCTC	2880
TTGCTTATGG TTTGGACAAG ACTGACAAAG AAGAAAAAAT CTTGGTATTT GACCTTGGTG	2940
GTGGTACATT CGACGTCTCT ATCCTTGAAT TGGGTGACGG TGTCTTCGAC GTATTGTCAA	3000
CTGCAGGGGA CAACAACTT GGTGGTGACG ACTTTGACCA AAAAATCATT GACCACTTGG	3060
TAGCAGAATT CAAGAAAGAA AACGGTATCG ACTTGTCTAC TGACAAGATG GCAATGCAAC	3120
GTTTGAAAGA TGCGGCTGAA AAAGCGAAGA AAGACCTTTC TGGTGTAAT TCAACACAAA	3180
TCAGCTTGCC ATTTATCACT GCAGGTGAGG CTGGACCTCT TCACTTGGAA ATGACTTTGA	3240
CTCGTGCGAA ATTTGACGAT TTGACTCGTG ACCTTGTGTA ACGTACAAAA GTTCCAGTTC	3300
GTCAAGCCCT TTCAGATGCA GGTTTGAGCT TGTCAGAAAT CGACGAAGTT ATCCTTGTG	3360
GTGGTTCAAC TCGTATCCCT GCCGTTGTTG AAGCTGTTAA AGCTGAACT GGTAAAGAAC	3420
CAACAAATC AGTAAACCT GATGAAGTAG TTGCTATGGG TGCGGCTATC CAAGGTGGTG	3480
TGATTACTGG TGATGTCAAG GACGTTGTCC TTCTTGATGT AACGCCATTG TCACTTGTA	3540
TCGAAACAAT GGGTGGAGTA TTTACAAAAC TTATCGATCG CAACACTACA ATCCCAACAT	3600
CTAAATCACA AGTCTTCTCA ACAGCAGCAG ACAACCAACC AGCCGTTGAT ATCCACGTT	3660
TTCAAGGTGA ACGCCCAATG GCAGCAGATA ACAAGACTCT TGGACGCTTC CAATTGACTG	3720
ATATCCCAGC TGCACCTCGT GGAATTCCTC AAATCGAAGT AACATTTGAC ATCGACAAGA	3780
ACGGTATCGT GTCTGTAAAG GCCAAAGACC TTGGAACCTA AAAAGAACAA ACTATTGTCA	3840
TCCAATCGAA CTCAGGTTTG ACTGACGAAG AAATCGACCG CATGATGAAA GATGCAGAAG	3900
CAACGCTGA AGCCGATAAG AAACGTAAAG AAGAAGTAGA CCTTCGTAAT GAAGTAGACC	3960

853

AAGCAATCTT TGCGACTGAA AAGACAATCA AGGAAACTGA AGGTAAAGGC TTCGACGCAG	4020
AACGTGACGC TGCCCAAGCT GCCCTTGATG ACCTTAAGAA AGCTCAAGAA GACAACAACT	4080
TGGACGACAT GAAAACAAAA CTTGAAGCAT TGAACGAAAA AGCTCAAGGA CTTGCTGTTA	4140
AACTCTACGA ACAAGCCGCA GCAGCGCAAC AAGCTCAAGA AGGAGCAGAA GGCGCACAAG	4200
CAACAGGGAA CGCAGGCGAT GACGTCGTAG ACGGAGAGTT TACGGAAAAG TAAGATGAGT	4260
GTATTGGATG AAGAGTATCT AAAAAATACA CGAAAAGTTT ATAATGATTT TTGTAATCAA	4320
GCTGATAACT ATAGAACATC AAAAGATTTT ATTGATAATA TTCCAATAGA ATATTTAGCT	4380
AGATATAGAG AATTATATTA GCTGAACATG ATAGTTGTAT CAAAAATGAT GAAGCGGTAA	4440
GGAATTTTGT TACCTCAGTA TTGTTGTCTG CATTGTATC GCGGATGGTA CCGTATCTGA	4500
CGAACGTTCA GCTTATAT	4518

(2) INFORMATION FOR SEQ ID NO: 122:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8145 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

TGCTATTTTC GATTCCCTTG GCGTTTTGA TTGCTTTGC CTTGCAAGTC CATTGGAAGC	60
CCCTCCATTA TCTGATTAAC ATTTACATCT GGGTTATGCG AGGAACCCCC TTA CTCTTGC	120
AACTGATTTT TATCTATTAT GTGCTCCCAA GTATTGGGAT TCGTTTAGAC CGCCTTCTTG	180
CAGCTATTAT TGCCTTTGTT CTCAACTATG CAGCTTACTT TGCAGAAATP TTCCGTGGGG	240
GAATTGACAC TATTCCAAGA GGACAGTATG AGGCCGCCAA GGTCTTGAAG TTTAGCCCTT	300
TTGACAGAGT GCGCTATATT ATCTTGCCCC AAGTGACCAA GATCGTTCTT CCTAGTGTCT	360
TTAATGAAGT TATGAGTTTG GTCAAGGATA CTTCTTTGGT CTATGCTCTC GGAATTTTCA	420
ACCTTATCTT GGCTAGTCGA ACAGCTGCTA ACCGCGATGC TAGTCTAGTT CCTATGTTCT	480
TGGCAGGAGC CATTTATTTG ATTTTGATTG GGATTGTGAC AATTATTTCC AAAAAAGTTG	540
AGAAGAAGTA TAGTTATTAT AGATAGGAGG CTGCCATGTT AGAATTACGA AATATCAATA	600
AAGTCTTTGG AGACAAACAA ATCCTGTCTA ATTTCACTCT AAGTATTCCT GAAAAGCAAA	660
TCCTGGCTAT CGTTGGACCT TCTGGTGGAG GTAAGACAAC TCTTTTACGT ATGCTTGCAG	720
GTCTTGAAC CATTGATTCA GGGCAAATCT TTTATAATGG ACAACCTTTA GAGCTGGATG	780

854

AATTGCAGAA GCGCAATCTA CTGGGATTTG TCTTCCAAGA TTTTCAACTA TTTCCTCATC	840
TATCAGTTCT GGAAAATTTG ACTTTATCGC CTGTGAAGAC CATGGGAATG AAGCAGGAAG	900
AGGCTGAGAA GAAGCCGAGT GGACTCTTGG AACAGTTAGG ACTAGGAGGA CACGCAGAGG	960
CCTATCCTTT CTCACTATCT GGTGGGCAAA AGCAGCGGGT GGCTTTGGCG CGTGCTATGA	1020
TGATTGACCC AGAAATCATT GGCTACGATG AACCAACTTC TGCCCTGGAT CCAGAATTAC	1080
GTTTGGAAGT GGAGAAGCTA ATCTTGCAAA ATAGGGAAC TGGGATGACC CAGATTGTGG	1140
TTACCCATGA TTTGCAGTTT GCTGAAAATA TCGCAGATGT ATTATTGAAA GTAGAACCTA	1200
AATAGGAGGA AAAATGGATG AAAAAATGGA TGCTTGATTT AGTCAGTCTG ATGACTGCTT	1260
TGTTCTTAGT AGCTTGTTGG AAAAAATCTA GCGAAACTAG TGGAGATAAT TGGTCAAAGT	1320
ACCAGTCTAA CAAGTCTATT ACTATTGGAT TTGATAGTAC TTTTGTTCCA ATGGGATTTG	1380
CTCAGAAAGA TGGTTCTTAT GCAGGATTTG ATATTGATTT AGCTACAGCT GTTTTTGAAA	1440
AATACGGAAT CACGGTAAAT TGGCAACCGA TTGATTGGGA TTTGAAAGAA GCTGAATTGA	1500
CAAAAGGAAC GATTGATCTG ATTTGGAATG GCTATTCCGC TACAGACGAA CGCCGTGAAA	1560
AGGTGGCTTT CAGTAACTCA TATATGAAGA ATGAGCAGGT ATTGGTTACG AAGAAATCAT	1620
CTGGTATCAC GACTGCAAAG GATATGACTG GAAAGACATT AGGAGCTCAA GCTGGTTCAT	1680
CTGGTTATGC GGACTTTGAA GCAAAATCCAG AAATTTTGAA GAATATTGTC GCTAATAAGG	1740
AAGCGAATCA ATACCAAACC TTAAATGAAG CCTTGATTGA TTTGAAAAAC GATCGAATTG	1800
ATGGTCTATT GATTGACCGT GTCTATGCAA ACTATTATTT AGAAGCAGAA GGTGTTTTAA	1860
ACGATTATAA TGTCTTTACA GTTGGACTAG AAACAGAAGC TTTTGCGGTT GGAGCCCGTA	1920
AGGAAGATAC AAAGTTGGTT AAGAAGATAA ATGAAGCTTT TTCTAGTCTT TACAAGGACG	1980
GCAAGTTCCA AGAAATCAGC CAAAAATGGT TTGGAGAAGA TGTAAGCAACC AAAGAAGTAA	2040
AAGAAGGACA GTAAGATAAA ATAGTGGCTG AAAGTGGCTT TTGATTAGCA AAACGTAGTT	2100
TTTTTTGTAA TCTAGGAAAA CGATAATAGC GATTGAATAT GGATAATTGA ATATGGAATA	2160
GCCCACTGTG ATTTCTAAAA CATTGTAAAA AATTGATTTG ACTTCCAAAA TTAAAATGTT	2220
CTGTAATGAA ATACTGATGT AACTGTTTGA GGAACAATAA AACGCATAAT ATCAAGGTTT	2280
TTGCACCTTA CATTATGCGT TTTTGTGATT TTAAGACTTG TTAGCTGATT TTTTACAATC	2340
CTGCGAAATC TTTGATTTCT TGTGCTGACA TTGAAGAGTC GCAACGGACG TTGATTTGTC	2400
CATCTGTAAT ATGAACAAAA CCTGGTACAG TTGGGATTCC ATAGCGTGAG CGGAATGCTT	2460
GCAAATCATT GAGTTGGCTT GGTCTTCAC TATTGATGAA GTAAATGTGA GCTTTGGTTT	2520
CAGCTACGAC ACCTGACAA GTACCTGCAA ATTTACGGCA GTAAGGGCAA GTTTTGGCAG	2580

855

CGATAAAGAA GGTTCAGTT TCTTTTTAT CAAGAGCTTC TTGCGCACGC ACAACTGTAG	2640
TGACTTCAAG GTCTTTGATG TTATCTAAAA ATTGTTCCAT GAGATTACCT CGCTTTCATT	2700
GATAAGTCTA GTATGCCATA AAGTTTCTAA AATTGCTTAG ATTTGATACG AAAAAAGATG	2760
AGGTTGGTTG GTCTCATCTT TTATAGGTCT TTATTTTACA AATGCATTGA TTTCTGCTTC	2820
GATGTTAGCA ATCTTAGCTT GTGATTCTTC GTTGGTTTCC CCTACAACG CAATGTAGAA	2880
CTTGATTTTT GGTCTGTAC CTGAAGGGCG AACGGCAATC CATGAACCGT CAGCAAGTGT	2940
GTATTTCAAC ACATCACTTG GAGGAGTTGT CAAGTTTGTA ACAGTACCGT CAGCAACAGT	3000
AGCAGTTTGT GCCTTGAAGT CTTCTACGAC AGTGATAGCT GTTGCCTTCC ATTCTGTTGG	3060
AGCATTGTTG CGGAATTTAG CCATAATCGC TTTGATTTGT TCAGCACCAT CGACACCTGA	3120
AAGAGTAACA GAGATTGTTT TTTCTGCGTA GTAGCCATAT TCTTTATAGA TTTCTTCGAT	3180
ACCGTCAGCA AGTGTCAAAC CACGAGAACG GTAGTAGGCA GCAAGTTCAG CAACTACAAG	3240
AACGGCTTGG ATGGCATCTT TATCACGTAC AAATGGTTTA ATCAAGTAAC CGAAGCTTTC	3300
TTCAAATCCC ATCATGTAAG TGTGGTGTG TTTTCTTCG AATCTTGGA TTTTTCAGC	3360
GATAAATTTG AAACCTGTCA AGACGTGAA CATAGTTGCG CCGTAGCTTT CAGCAATCTT	3420
CGTTACCAAG TCAGTTGAAA CGATAGATTT GCAGAGAGCG GCATTTTCAG GAAGAGTTCC	3480
AGCGTTTTTG TGAGCTTCCA AGATGTATTT AGCCATGATA GCACCGATTT GGTACCTGA	3540
AAGGTTGAGG TAGCTACCAT CTTTTGAAG AACTTCAACA CCAACACGGT CAGCGTCTGG	3600
GTCAGTTGCG ACAAGAACAT CTGCACCAAC TTGACGACCA AGTCTTCAG CAAGGGCAAA	3660
GGCTGCTTGG CTTTCTGGGT TTGGAGATGT TACAGTTGAA AAGTCTGGGT CAGCAGTTGC	3720
TTGCGCTTCA ACAACTTGAA CAGAGTCAA TCCTGCTTGG GCAAGAGCAC GACGAGCCAA	3780
CATTTACCA GTACCATGAA GTGGTGTGTA GACAATCTTC ATGTCTTTAC CAAATCTTC	3840
AATCAAGGCT GGGTTGATGT TTATGTCCTT AACCTCTTTA AGGTATTCTA TGTCAACAGC	3900
TTGCGCGATA ACTTCAATCA AGCCAGAAGC TTTTTCAGTT TCCACATCAG CAACTTCAAC	3960
TGCAAATGGG TTTTCGATTG CACGGATATA AGTAGTCAA GCGTCCGCAT CGTGTGGAGG	4020
CATTTGTCCA CCGTCTTCAC CGTAAACCTT GTAACCGTTA AATGGAGCAG GGTGTGGCT	4080
GGCTGTGACC ATGATACCTG CGAAACAGTT GAGATGACGA ACTGCAAATG ATAGTTCTGG	4140
AGTCGGACGA AGGCTTTCAA ATACGTAAGA TTTGATGCCG TGTTTAGCAA GAACTGCCGC	4200
AGATTCAAAG GCAAACCTCAG GTGAGAAGTG ACGGCTATCG TAGGCAATTG CTACACCGCG	4260
TTCTTTCTCG TTTCCACCTT TTGACTCAAT CAAACGAGCC AATCCTTCAG TAGCTTGCG	4320

856

AACAACGTAG ATGTTGATAC GGTTTGTACC AGCACCAACC AAGCCACGCA TACCTGCAGT	4380
ACCAAATTCA AGATTGTAT AGAAGGCATC TTCCTTAGTT TTTTCGTCCA TATTTTCCAA	4440
ATCTTGACGA AGGTAGTCAC GAAGCTCCAC AAAATCAACC CATTCTCTGGT AATTTTCTTG	4500
GTAAGACATT CAAATTCTCC TTTATTTTAA AAACATTTAA TCAGTTTAAT TATATCATTT	4560
TTTTTAGTTT TAGTAAAACC TTATCTGCTT CGAACATCTC TTCAAACCAG GTCAGATTGA	4620
ATTTTGGGGT TATATGATGT TGAGGCTAGG AAAAAATCAA TTTCAGTAAA AAAAGTAAGT	4680
CTTCTCATAA CAAAACATTG ATATAGTTAC TTAGTTTTAA ACAAGCATAT TATAATAAAG	4740
CTATGGCATA TAGTACTGAT TTTAAACAGC GAGCATTAGA TTACATCAAA GAGGGGCACA	4800
GCCATGTCGA GGCAGCCAAG TTTTTTGGTG TTGGCGTCAG AACTCTCTTC ACGTGGGAAA	4860
AGAAAGACGT GAACAAGAAC ACATAGAGAG GAAAAAGCGA GTCGTCAAAA ACCGAAAGAT	4920
TCCTTTAGAG GAATTGAAAG CCTTTGTAGA GGTCATCCA GATGCTTTTT TACGGGAAAT	4980
TGCGGCACAT TTTGATTGTG CTGTTCCCTC AGTATGGGCA GCTTTAAAGC AGATTAAGGT	5040
CACTTTAAAA AAAGATGACG AGCTTTAAGG AACAAGACCC AGAAAAGTAG CCTTATTTCT	5100
TAAGAATTTT AATAGTTTAA AGCACCTAGC ACCTGTTTAT ATTGATGAAA CAGGAATCGA	5160
CCGCTATCTC TATCGTCCTT ATGCAGGGGC TCCTAGAGGG GAGAAAGTCT ATGAAAAGAT	5220
TAGCGGACGT CGTTTTGAGC GAACTTCAAT TGTTCAGGA CAAGTAGACG GAGAGTTTAT	5280
AGCTCCCATG ATTTACAAGA AAAGCATGAC AAGCGATTTC TTTGTGGAGT GGTTCAAAAC	5340
GCAACTCCTA CCTGCTTTGA AGACACCTCA TGTATTGTG ATGGGCAATG CTGCTTTTCA	5400
TCCAAGAAC ATTTTGGATG AACTCTGCAT CCAAGATAAA CACTTTTCT TACCTCTACC	5460
ACCTTATCA CCGGATTGTA ATCCTATTGA GCAAGCTTGG GCTATCTTGA AAAAGAAAGT	5520
GACGGATGTA TTAAGGGAAG TTCCAATAT TTTGAATGT TTGGAATGCT TTTTAAAAAC	5580
TAGATGACTA TAACGGTTCT AAAGGAACCT ATCGAGTAGT CATTAAAACT AAGGATACTG	5640
CTGGTTAAGA GAAGACGTA TACAATCAAA CCATTACCG TGTAGCCGAA ATCGTTCAGA	5700
ATGAAGACTT GTATCAGAAT GAAGACTTGT ATAAGAAAGG TTTGAATGTT GAACTTCCGC	5760
ACCAACAAAT TAAGGGATTT TTTGAAGCAG AGTTTAAAAA TCGTATTAA TGGAGTTCTTA	5820
ATACTAAAAT AAAAAATAGT ACATTAAATC GTGTAAATAA AAAAACTATA CACCAGAGCA	5880
ACAAAAACTC CATGATCAAT TTGAAGCAGA AGCAACGGAA GATGCTAAAA AACAAGGCCA	5940
TATTGTGTTG AATGTTGACC AGGATTTCAT GAGCATATCT AAGTCTAATA AAAGTGGTTC	6000
AGACTGGAAG AAAACTTTCA CAGTGAGGAT AACCAATAGG CTAGCAAATG ACTTGAATAA	6060
TGTCTTGAAA CAGGTTGATA AAGATACTCC TAATACCCCA ACTTGGCTAA ACTCAGCTCC	6120



857

TTCTAAAGCT AAAGATGATG ACAGACTATA TAAACTACTG AAGACTCTTA TACCAGGAGA	6180
AAATTACCTA TCATGTTAAG GATAATCAGC TAGAAGTAGA AACAGATAAA TACACATATA	6240
CTGCCGCTAG AAATGGTAGT AAGGAAGTTG GTATTCAAGA GTCAGATATA GCAGCAACTC	6300
TAAGTGCCGA TGAATATAAT TCTAATCGCC AAACCTTTTGA GAGAGAATAC AAATACAAAA	6360
GCAAAATGCCC TTAATAATGG TTGGGCTAGA TCTGGTTCTG AAGAGTTCAA AAAGTTCTCC	6420
CACTTTGTAG GGGTAGACAA AGGGATTGTG CGAACGAATG TACTGACTGG TAAAAAACTA	6480
TCTGATAAGA TTAGGAAAGA AGTGGGCTCT GGAGATAGCA AACTAGGAAA AGGCGGTAT	6540
TTCTCTACTG GGGATGTTCT ATTAGGAAAA GATGTTGTTT CTTATACCGT ACAAGTATTT	6600
TCAGAGAATA ATGAAAGAGT AGGAGTAAAC ACTCAAAGTC ACCGTGTCA GTATAATCTC	6660
CCAATTCTAG CTGACTTTTC AGTCATCCAA GATACTGTGG AACCATCACG AACCCTGTGT	6720
GAAAAAATCA TTCCAAAACCT AAATATTCCC GAAGAAGAGA AAGGGAAAAAT AACCGAAGAA	6780
ATCAAGAAAA AGAAAAAAC CTCAGAATTG GCAGAACTAA TCTCAGAAAA TGTGAAAGTT	6840
CGCTATGTTG ATGAACAAGG GCGTTTGCTA TCATTGAAAA ATGATACTGG AATTGGAGAA	6900
AAAGAAAGTG ACGGAACCTA CATTACCAAT AAAAAACAAC TGATTGGTAC CAGCTATAAT	6960
GTCACAGATA AAAAAGCTAG TAGCATGACT ACTACTGACG GAAAAATTA TACTTTTAAA	7020
GAAGCAGATA CAAATTCTGC AAGTTTAACT GGGAAATATTG TAAGCGAAGG TAGAACAGTG	7080
ACCTTAGTTT ATAGAGAAAG CGAAGCGCCA ACCACTGCTA CAGTAACAGC CAATTACTAT	7140
AAAGAAGGTA GGCAAGAGAA GTTGGTAGAG TCTGTTATAA AAGCTGATTT AGCGATAGGT	7200
TCTGAGTATA CCACAGAATC AAAAAGTATT GAAGGGAAAA CAACAACTGA GGACAAAGAA	7260
GACCGAGTTA TCACAAGGAA AACAACATAC ACCTTGCTAG CAACTCCTGA AAATGCGTAC	7320
CAGAAGACGG TGCAACAGTT GACTATTACT ACCGTGAGAA TGTGAGGAA ACAGTGGTTC	7380
CCAAAACAGC AACCTCTACT GAGACGAAGA CTATAACGCG TATCATTCTAT TACGTTGATA	7440
AAGTTACGAA CCAAAATGTA AAAGAAGATG TTGTTCAACC TGTAACCTTA AGCCGTACAA	7500
AAACTGAGAA CAAGGTCACG GGAGTTGTAA CCTACGGTGA ATGGACAACA GGAAACTGGG	7560
ACGAGGTTAT ATCTGGTAAG ATTGACAAGT ACAAAGATCC AGATATTCCA ACAGTTGAAT	7620
CACAAGAAGT TACGTGAGAC TCTAGTGATA AAGAAATAAC GGTAAGGTAT GACCGTTTAT	7680
CAACACCAGA AAAACCAATC CCACAACCAA ATCCAGAGCA TCCAAGTGTT CCGACACCAA	7740
ACCCAGAACT ACCAAATCAA GAGACTCCAA CACCAGATAA ACCAACTCCA GAACCAGGTA	7800
CTCCAAAAAC TGAAACTCCA GTGAATCCAG ACCCAGAAGT TCCGACTTAT GAGACAGGTA	7860

858

AGAGAGAGGA ATTGCCAAAC ACAGGTACAG AAGCTAATGC TACCTTGGCT AGTGCTGGTA	7920
TCATGACCTT GTTAGCTGGT CTAGGATTAG GATTTTTCAA GAAAAAAGAA GATGAAAAAT	7980
AATAGATTTT AGAATCTAGG AACCAGGAAA AGCTCACAGA TGTGGGCTTT TTTCCTGGTT	8040
TTGAGAACGA GGTCTTTCGT AAAGAATAAA AACGCTTACA AGTCTGTTGA ACTGGGAAAC	8100
TATGAATCCT ATTTTTTTAA AAATATTTC AGAAATCAGT TGCGG	8145

(2) INFORMATION FOR SEQ ID NO: 123:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8697 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

CGGTACCGGG AACGATACTT AGTCTAATTT TGCACCTTTT CCATGTATGG TAAAGGTTTT	60
TCTTTTTTTA AAAAGGAAAA CGAGAAGAGG AGGTTCTTAT GAAAGCAAGC ATTGCCTTGC	120
AAGTTTTACC CCTAGTACAG GGGATTGATC GGATAGCTGT TATTGATCAG GTCATTGCTT	180
ATCTGCAWAC TCAAGAAGTG ACGATGGTAG TGACACCATT TGAAACGGTC TTGGAAGGGG	240
AGTTTGATGA GCTTATGCGC ATTCTAAAAG AAGCGCTGGA AGTGGCAGGG CAGGAGGCAG	300
ACAATGTCTT TGCCAATGTC AAAATAAATG TAGGAGAGAT TTAAAGTATT GATGAGAAAC	360
TTGAGAAGTA TACTGAGACG ACACATTAGT CTATTGGGCT TTCTCGGAGT ATTGTCAATC	420
TGGCAGTTAG CAGGTTTTCT TAAACTTCTC CCCAAGTTTA TCCTGCCGAC ACCTCTTGAA	480
ATTCTCCAGC CCTTTGTTCTG TGACAGAGAA TTTCTCTGGC ACCATAGCTG GGCGACCTTG	540
AGAGTGGCTT TACTGGGGCT GATTTTGGGA GTTTTGATTG CCTGTCTTAT GGCTGTGCTC	600
ATGGATAGTT TGACTTGGCT CAATGACCTG ATTTACCCTA TGATGGTGGT CATTGAGACC	660
ATTCCGACCA TTGCCATAGC TCCTATCCTG GTCTTGTGGC TAGGTTATGG GATTTTGCCC	720
AAGATTGTCT TGATTATCTT AACGACAACC TTTCCCATCA TCGTTAGTAT TTTGGACGGT	780
TTTAGGCATT GCGACAAGGA TATGCTGACC TTGTTTAGTC TGATGCGGGC CAAGCCTTGG	840
CAAATCCTGT GGCATTTTAA AATCCCAGTT AGCCTGCCTT ACTTTTATGC AGGTCTGAGG	900
GTCAGTGTCT CCTACGCCCTT TATCACAAC TGGGTATCTG AGTGGTTGGG AGGTTTTGAA	960
GGTCTTGGTG TTTATATGAT TCAGTCTAAA AAAGTGTTC AGTATGATAC CATGTTTGCC	1020
ATTATTATTC TGGTGTGAT TATCAGTCTT TTGGGTATGA AGCTGGTCGA TATCAGTGAA	1080
AAATATGTGA TTAATGGAA ACGTTCGTAG AATTAGAATG TTTCTGAAAA AGAAAAGAGG	1140

859

AAATCAAAAT GAAGAAAACA TGGAAAGTGT TTTTAACGCT TGTAACAGCT CTTGTAGCTG	1200
TTGTGCTTGT GGCCTGTGGT CAAGGAACTG CTTCTAAAGA CAACAAAGAG GCAGAACTTA	1260
AGAAGGTTGA CTTTATCCTA GACTGGACAC CAAATACCAA CCACACAGGG CTTTATGTTG	1320
CCAAGGAAAA AGGTTATTTC AAAGAAGCTG GAGTGGATGT TGATTGAAA TTGCCACCAG	1380
AAGAAAGTTC TTCTGACTTG GTTATCAACG GAAAGGCACC ATTTGCAGTG TATTTCCAAG	1440
ACTACATGGC TAAGAAATTG GAAAAAGGAG CAGGAATCAC TGCCGTTGCA GCTATTGTTG	1500
AACACAATAC ATCAGGAATC ATCTCTCGTA AATCTGATAA TGTAAGCAGT CCAAAAGACT	1560
TGGTTGGTAA GAAATATGGG ACATGGAATG ACCCAACTGA ACTTGCTATG TTGAAAACCT	1620
TGGTAGAATC TCAAGGTGGA GACTTTGAGA AGGTTGAAAA AGTACCAAAT AACGACTCAA	1680
ACTCAATCAC ACCGATTGCC AATGGCGTCT TTGATACTGC TTGGATTAC TACGGTTGGG	1740
ATGGTATCCT TGCTAAATCT CAAGGTGTAG ATGCTAACTT CATGTACTTG AAAGACTATG	1800
TCAAGGAGTT TGACTACTAT TCACCAGTTA TCATCGCAAA CAACGACTAT CTGAAAGATA	1860
ACAAAGAAGA AGCTCGCAAA GTCATCCAAG CCATCAAAAA AGGCTACCAA TATGCCATGG	1920
AACATCCAGA AGAAGCTGCA GATATTCTCA TCAAGAATGC ACCTGAACTC AAGGAAAAAC	1980
GTGACTTTGT CATCGAATCT CAAAAATACT TGTCAAAGA ATACGCAAGC GACAAGGAAA	2040
AATGGGGTCA ATTTGACGCA GCTCGCTGGA ATGCTTTCTA CAAATGGGAT AAAGAAAATG	2100
GTATCCTTAA AGAAGACTTG ACAGACAAAG GCTTCACCAA CGAATTTGTG AAATAATGAC	2160
AGAAATTAGA CTAGAGCAGC TCAGTTATGC CTATGGTCAG GAGAGGATTT TAGAGGATAT	2220
CAACCTACAG GTGACTTCAG GCGAAGTGGT TTCCATCCTA GGCCCAAGTG GTGTTGAAAA	2280
GACCACCCTC TTAAATCTAA TCGCTGGGAT TTTAGAAGTT CAGTCAGGGA GAATTGTCCT	2340
TGATGGTGAA GAAAATCCCA AGGGGCGCGT GAGTTATATG TTGCAAAGG ATCTGCTCTT	2400
GGAGCACAAG ACGGTGCTTG GAAATATCAT TCTGCCCTC TTGATTCAA AGGTGGATAA	2460
GGCAGAAGCT ATTTCCCGAG CGGATAAAAT TCTTGCGACC TTCCAGCTGA CAGCTGTAAG	2520
AGACAAGTAT CCTCATGAAC TTAGCGGTGG GATGCGCCAG CGTGTAGCCT TACTCCGGAC	2580
CTACCTTTTT GGGCACAAGC TCTTTCTCTT AGATGAGGCC TTTAGCGCCT TGGATGAGAT	2640
GACAAAGATG GAACTCCAGC CTTGGTATCT TGAGATTAC AAGCAGTTGC AGCTAACAAC	2700
CCTGATCATC ACGCATAGTA TTGAGGAGGC CCTCAATCTC AGCGACCGTA TCTATATCTT	2760
GAAAAATCGC CCTGGGCAGA TTGTTTCAGA AATTAACTA GATTGGTCTG AAGATGAGGA	2820
CAAGGAAGTC CAAAAGATTG CCTACAAACG TCAAATTTTG GCGGAATTAG GCTTAGATAA	2880

860

GTAGAAAAAT AGGGAGTTGG TGAAGATTAT CCTTTACCAG CGCCCTTTT CTTTAAAAA	2940
TGAGAAAATT TCGGTATAAT AGTCAAACAA GGTCAAGGTT TAAAGAGAGA GGTGGGTTTG	3000
TTATGAGATT TAAAAATACA TCGGATCATA TTGAGGCCTA CATCAAGGCG ATTTTAGATC	3060
AATCTGGTAT CGTGGAGTTG CAACGGAGTC AGTTGGCAGA TACCTTTCAG GTTGTTCCTA	3120
GTCAGATTAA CTACGTGATC AAGACACGCT TTACGGAAAG TAGAGGCTAC TTGGTTGAAA	3180
GTAAGCGTGG TGGCGGAGGC TACATTCTGA TAGGACGGAT TGAGTTTTCT AGTCATCATG	3240
AAATGCTCCG GGAGCTGCTT TACTCGATTG GTGAGCGAGT CAGTCAAGAA ATTTATGAGG	3300
ATATTCTCCA GCTTTTGGTT GAGCAGGAAT TGATGACCAA GCAGGAGATG AATTGCTAG	3360
AATCAGTAGC TTTGGATCGC GTTTTAGGAG AAGAAGCTCC AGTTGTTCTGA GCAAACATGC	3420
TACGTCAGAT CATAAAGAG GTAGATAGAA AAGGGAAGTA AGATGAACTA TTCAAAAGCA	3480
TTGAATGAAT GTATCGAAAG TGCCTACATG GTTGCTGGAC ATTTTGGAGC TCGTTATCTA	3540
GAGTCGTGGC ACTTGTTGAT TGCCATGTCT AATCACAGTT ATAGTGTAGC AGGGGCAACT	3600
TTAAATGATT ATCCGTATGA GATGGACCGT TTAGAAGAGG TGGCTTTGGA ACTGACTGAA	3660
ACGGACTATA GCCAGGATGA AACCTTTACG GAATTGCCGT TCTCCCGTCG TTTGCAGGTT	3720
CTTTTGTATG AAGCAGAGTA TGTAGCGTCA GTGGTCCATG CTAAGGTA CTAGGGACAGAG	3780
CACGTCCTCT ATGCGATTTT GCATGATAGC AATGCCTTGG CGACTCGTAT CTTGGAGAGG	3840
GCTGGTTTTT CTTATGAAGA CAAGAAAGAT CAGGTCAAGA TTGCTGCTCT TCGTCGAAAT	3900
TTAGAAGAAC GGGCAGGCTG GACTCGTGAA GATCTCAAGG CTTTACGCCA ACGCCATCGT	3960
ACAGTAGCTG ACAAGCAAAA TTCTATGGCC AATATGATGG GCATGCCGCA GACTCCTAGT	4020
GGTGGTCTCG AGGATTATAC GCATGATTTG ACAGAGCAAG CGCGTTCTGG CAAGTTAGAA	4080
CCAGTCATCG GTCGGGACAA GGAAATCTCA CGTATGATTC AAATCTTGAG CCGGAAGACT	4140
AAGAACAACC CTGCTTTGGT TGGGGATGCT GGTGTCGGGA AAACAGCTCT GGCGCTTGGT	4200
CTTGCCACGC GTATTGCTAG TGGTGACGTG CCTGCGGAAA TGGCTAAGAT GCGCGTGTTA	4260
GAACCTTGATT TGATGAATGT CGTTGCAGGG ACACGCTTCC GTGGTGACTT TGAAGAACGC	4320
ATGAATAATA TCATCAAGGA TATTGAAGAA GATGGCCAAG TCATCCTCTT TATCGATGAA	4380
CTCCACACCA TCATGGGTTT TGGTAGCGGG ATTGATTCGA CTCTGGATGC GGCCAAATATC	4440
TTGAAACCAG CCTTGGCGCG TGGAACCTTG AGAACGGTTG GTGCCACTAC TCAGGAAGAA	4500
TATCAAAAAC ATATCGAAAA AGATGCGGCA CTTTCTCGTC GTTTCGCTAA AGTGACGATT	4560
GAAGAACCAA GTGTGGCAGA TAGTATGACT ATTTTACAAG GTTTGAAGGC GACTTATGAG	4620
AAACATCACC GTGTACAAAT CACAGATGAA GCGGTTGAAA CAGCGTTAA GATGGCTCAT	4680

861

CGTTATTTAA CCAGTCGTCA CTTGCCAGAC TCTGCTATCG ATCTCTTGGA TGAGGCGGCA	4740
GCAACAGTGC AAAATAAGGC AAAGCATGTA AAAGCAGACG ATTCAGATTT GAGTCCAGCT	4800
GACAAGGCCC TGATGGATGG CAAGTGGAAG CAGGCAGCCC AGCTAATCGC AAAAGAAGAG	4860
GAAGTACCTG TCTACAAAGA CTTGGTGACA GAGTCTGATA TTTTGACCAC CTTGAGTCGC	4920
TTGTCAAGAA TCCCAGTTCA AAAACTGACT CAAACGGATG CTAAGAAGTA TTTAAATCTT	4980
GAAGCAGAAC TCCATAAAGG GGTATCGGT CAAGATCAAG CTGTTTCAAG CATTAGCCGT	5040
GCCATTCGCC GCAACCAGTC AGGGATTGCG AGTCATAAGC GTCCGATTGG TTCCTTTATG	5100
TTCTAGGGC CTACAGGTGT CGGGAAAAC GAATTAGCCA AGGCTCTGGC AGAAGTTCTT	5160
TTTGACGACG AATCAGCCCT TATCCGCTTT GATATGAGTG AGTATATGGA GAAATTTGCA	5220
GCTAGTCGTC TCAACGGAGC TCCTCCAGGC TATGTAGGAT ATGAAGAAGG TGGGGAGTTG	5280
ACAGAGAAGG TTCGCAATAA ACCCTATTCC GTTCTCCTCT TTGATGAGGT AGAGAAGGCC	5340
CACCCAGATA TCTTTAATGT TCTCTGCAG GTTCTGGATG ACGGTGTCTT GACAGATAGC	5400
AAGGGACGCA AGGTCGATTT TTCAAATACC ATTATCATTG TGACATCGAA TCTAGGTGCG	5460
ACTGCCCTTC GTGATGATAA GACTGTTGGT TTTGGGGCTA AGGATATTCG TTTTGACCAG	5520
GAAATATGG AAAAACGCAT GTTTGAAGAA CTGAAAAAG CTTATAGACC GGAATTCATC	5580
AACCGTATTG ATGAGAAGGT GGTCTTCCAT AGCCTATCTA GTGATCATAT GCAGGAAGTG	5640
GTGAAGATTA TGGTCAAGCC TTTAGTGGCA AGTTTGACTG AAAAAGGCAT TGACTTGAAA	5700
TTACAAGCTT CAGCTCTGAA ATTGTTAGCA AATCAAGGAT ATGACCCAGA GATGGGAGCT	5760
CGCCCACTTC GCAGAACCCT GCAAACAGAA GTGGAGGACA AGTTGGCAGA ACTTCTTCTC	5820
AAGGGAGATT TAGTGGCAGG CAGCACACTT AAGATTGGTG TCAAAGCAGG CCAGTTAAAA	5880
TTTGATATTG CATAAAAGAA TAAAAGTATC AGCATCTGAC CATAAGTCAC AGTGGAGTGA	5940
AATTCAATGA AAATCAAAGA GCAAAC TAGG CAGCTAGCCG CAGGTTGCTC AAAACACTGG	6000
TTTGAGGTTG CAGATAGAGC TGACGTGGTT TGAAGAGATT TTCGAAGAGT ATGAAACTAA	6060
AACCTATAGC TTCTAAACGA TCCGTGGTTT TCATCATTCA ACACAAAATT CATATGTTTA	6120
TTACCCCTCG TCGTATTGT CTTAGAGCGT GTGTAGTAGA AAAAGAGCAG TCTTATCTGA	6180
AATTTTATT CTTTCAAAG AGACCTGTTT CTTTTTGCA TGTCAAATCC GTTCTAGCTG	6240
GTATTTGAAA AATCAAATA ATATTCAATG AAAATCAAAG AACAACTAG GAAGCTAGCC	6300
GCAGGTGCT CAAAACACTG TTTTGAGGTT GTAGATAGAG CTGACGTGGT TTGAAGAGAT	6360
TTTCGAAGAG TATAAGCTGC AAGATGAATG ATTTTCTTGT ATTGACGTTG TTGTTGACAA	6420

862

AAAGTAGCGG ATAAATGAAA TCCATTCCAT TATCATAGAT GATAGGCTGG TAGGAAATTT	6480
TCAAATAGCA TACAGGAAAT AGATGTATGG AGTTCTGGTA GTAGAAAGGG AGAGAGATGA	6540
ACATTTTAGT TGCAGATGAC GAGGAAATGA TTAGAGAAGG AATTGCAGCA TTTCTGACAG	6600
AAGAGGGTTA TCATGTCATT ATGGCTAAGG ATGGACAAGA GGTCTTGGA AAATTTCAAG	6660
ATCTCCCTAT CCATCTCATG GTACTGGATT TAATGATGCC TAGGAAGAGT GGTTTTGAAG	6720
TGTTAAAAGA AATCAATCAA AAGCACGATA TTCCTGTCAT CGTCTTGAGT GCTCTGGGAG	6780
ATGAAACTAC TCAGTCACAG GTATTTGATC TCTATGCTGA TGATCATGTG ACAAACCTT	6840
TTTCTTTGGT ACTGCTTGTC AAGCGTATTA AGGCGCTTAT CAGACGTTAC TACGTCATAG	6900
AGGATCTTTG GCGATATCAG GATGTAACAG TGGATTTTAC CTCTTACAAA GCACATTATA	6960
AAAATGAAGA AATTGATCTC AAACCAAAGG AATTACTGGT ACTAAAGTGT TTGATTTCAGC	7020
ATAAAAATCA AGTTTAAAGT AGAGAGCAGA TATTGGAAGA AATTTCAAAA GATGTAGCTG	7080
ATTTACCTTG TGATAGGGTC GTTGATGTCT ATATTCGTAC TCTTCGCAA AAATTAGCTT	7140
TAGATTGTAT CGTGA CTGTG AAAAATGTTG GGTATAAGAT TAGCTTATGA TAAAAATCC	7200
TAAATTATTA ACCAAGTCTT TTTAAGAAG TTTTGCAATT CTAGGTGGTG TTGGTCTAGT	7260
CATTCATATA GCTATTTATT TGACCTTTCC TTTTATTAT ATTCAACTGG AGGGGGAAAA	7320
GTTTAATGAG AGCGCAAGAG TGTTTACGGA GTATTTAAAG ACTAAGACAT CTGATGAAAT	7380
TCCAAGCTTA CTCCAGTCTT ATTCAAAGTC CTTGACCATA TCTGCTCACC TTAAGAGAGA	7440
TATTGTAGAT AAGCGGCTCC CTCTTGTCGA TGA CTGGAT ATTAAGATG GAAAGCTATC	7500
AAATATATC GTGATGTTAG ATATGTCTGT TAGTACAGCA GATGGTAAAC AGGTAACCGT	7560
GCAATTTGTT CACGGGGTGG ATGTCTACAA AGAAGCAAAG AATATTTTGC TTTTGTATCT	7620
CCCATATACA TTTTGGTTA CAATTGCTTT TTCCTTTGTT TTTTCTTATT TTTATACTAA	7680
ACGCTTGCTC AATCCTCTTT TTTACATTTC AGAAGTGA CTGCTAAATGC AAGATTGGA	7740
TGACAATATT CGTTTGTATG AAAGTAGGAA AGATGAAGTT GGTGAAGTTG GAAAACAGAT	7800
TAATGGTATG TATGAGCACT TGTGAAGGT TATTTATGAG TTGGAAAGTC GTAATGAGCA	7860
AATTGTAAAA TTGCAAAATC AAAAGGTTTC CTTTGTCCGC GGAGCATCAC ATGAGTTGAA	7920
AACCCCTTTA GCCAGTCTTA GAATTATCCT AGAGAATATG CAGCATAATA TTGGAGATTA	7980
CAAAGATCAT CCAAAATATA TTGCAAAGAG TATAAATAAG ATTGACCAGA TGAGCCACTT	8040
ATTAGAAGAA GTACTGGAGT CTCTAAATT CCAAGAGTGG ACAGAGTCTC GTGAGACCTT	8100
GACTGTAAAG CCAGTTTATG TAGATATTTT ATCACGTTAT CAAGAATTAG CTCATTCAAT	8160
AGGTGTTACA ATTGAAAATC AATTGACAGA TGCTACCAGG GTCGTCATGA GTCTTAGGGC	8220

863

ATTGGATAAG GTTTGTGACAA ACCTGATTAG TAATGCAATT AAATATTCAG ATAAAAATGG	8280
GCGTGTAATC ATATCCGAGC AAGATGGCTA TCTCTCTATC AAAAATACAT GTGCCCTCT	8340
AAGTGACCAA GAACTAGAAC ATTTATTTGA TATATTCTAT CATTCTCAA TCGTGACAGA	8400
TAAGGATGAA AGTCCCGGTT TGGGTCTTGA CATTGTGAAT AATATTTTAG AAAGCTATCA	8460
AATGGATTAT AGTTTTCTCC CTTATGAACA CGGTATGGAA TTTAAGATTA GCTGTAGAC	8520
AGATTAGTTT TTTATTAAAG TTCATATAGG GTTAACATAA GTGTGTTATT CTTTGTGTAG	8580
ATAAAAGAAA GGATACTAAT ATGGTATTAG CGATTATTTT AGTAACATTC TTTATTCGAT	8640
TGATTTTTTTT AAAGCGTTTC ATAGAGAATG AGAAACGAAT CCTTAGCAAT GGCGGGG	8697

(2) INFORMATION FOR SEQ ID NO: 124:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4317 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

AACCATACAT ACGGCAAGGC AAAGCTGACG CGGTTTGAAG AGATTTTCGA AGAGTATTAG	60
TTGCCTTTAA AGGCATCCAC CATCGTTTGA AATTCTTCAT TTGAGAGAGT AATCCCTTTG	120
CCCATTTTAG TATGGTCTGG ACTCCAAGCA CGAATATCAA ACTTTGCAGG GGCACCATTA	180
AAGCTCACAC GGTTAATTTC CTTGGTCCAA CCTTTTTCGT TTTCAGAAAG AGTCAACAAG	240
TGCTCTTCGA TTTCAAATGT AAATTCTGCC ATTTTCTTCT CCTTTTTCAG TTTTCATTAGT	300
TTATTTCGTAA AATCTTGTAG ATTTTAGGAA AATTTTATAT AATATTGATA TAAAAGAAGG	360
GAGGCCAATA TGAGACATAA ATTCCAGCAA GTTCTAAATA AAATACATGA TTTTAAAT	420
GGATATGACC AACCTGACCA GACTGAAACC AACTCCCTTA CAGCCACTAT TGAAGAGGCT	480
ATCCAGAAAC AAACCGCTGT TCACCTTATC TTGTCTGAGA CAAGCTTTAC AGGTGACATC	540
ATCAAATATG ATCAGCAAGG CCAGCAAATT ATCGTGAAAA ATTTTTCCAA AAATGTGAGC	600
CGGATTATCC GTATAAGCGA TATTCAACGC CTGCGATTG TCCCTCAAC TGTCCAAACA	660
GCCCCAAAAA ATAGATTTAA GAAAGAGTGA GATGTAGTTG CTTTCATCCA CTCTTTTTC	720
TTAGCGAATT TGTTCAAAAT GTAAATGAAC TGCGATATGA TCTCCATAAC CACTTCTTTC	780
CAAGTCACGT TGTAACGAT AGGAAATGTA GTGTCTGCA ATGGTAATGT AACCTGCGCC	840
CAATAAACGA TGTTCACCA TAGATTGAAT CATACTGATA GTCGCACGTT CCACCTTGGC	900

864

TTCTTGTA	AACTCCAA	CTTCTTAGT	GACTTGAGCA	AGATTTTGAC	GCAAATCATC	960
TGTCAAAACA	TAAACAGTTT	GGGCTGCCTT	CAAGATGGCT	TGGTAAATCT	TATCTGGATT	1020
AAATTCAGCA	ATTCGCCAT	TACGTTTGAT	TACTTGACATA	GGTTTCTCCT	TTATTCTTTG	1080
TTTCTTTGA	TTTCTGCCAG	CATTTTTTCT	TCTTCTACTG	TCAGTTGATA	ATGTTCAAGT	1140
AAATCCGGTC	TGCGCTCGTA	GGTTTTCTTT	AAACTCTCGT	ACAATCGCCA	CTGACGAATC	1200
TTTTTCATGGT	GGCCACTCAT	CAATACATCT	GGCACGACCA	TGCCTCGATA	ATCATAGGGA	1260
CGTGTGTACT	GAGGATATTC	TAAAAGACCT	GAAGAAAAAC	TATCATCTTG	GTGGCTAGAC	1320
TCCTTGCCAA	TCACCTCTGG	AATCAGGCGA	ACTGTAGCAT	CAATCATGGT	CATAGCTGCC	1380
AATTCTCCAC	CAGTGAGGAC	ATAGTCACCT	AGGGAAATCT	CATCTGTAC	CAAGGTCTTA	1440
ATGCGCTCAT	CATAACCCCT	ATAGTGCCCA	CAGATAAAGA	TTAGCTCTTC	CTCTTGAGCC	1500
AAATCTTCAG	CATAAGCCTG	ATCAAAGTGC	TTTCCAGCAG	GATCAAGGAG	AATAACGCGC	1560
GGATTTTTCT	TTTCAATAGC	ATCAAAGGAA	TCGAAAATAG	GTTGTGCTCT	GAGCAACATG	1620
CCCTGACCGC	CTCCGTAGGG	CTCATCATCT	ACATGACGGG	CCTTTTCAGC	ATTTTCTCGA	1680
AAATTATGAT	ACTGGATATC	CAAGAGCCCT	TTTTCTCGAG	CCTTTCCAAC	GATTGAGTGC	1740
TCCAGTGGAG	AAAACATCTC	TGGAAGAGG	GTTAAATAT	CAATCTTCAT	CGTCTAACCC	1800
TTCTAAGATT	TCCACATCGA	CCCGTTTACT	TGGAATATCA	ACATTGAGAA	CCACTGGTGG	1860
GATATAAGGT	AAAAGCAAAT	CACGTTTGCC	TTTTCGTTTG	ACCACCCAGA	CATCATTAGC	1920
ACCTGGTTGC	AGGATTTCCT	TGATGGTTCC	AACCAAGCTA	TCACCCCTCAT	AGACTTCCAA	1980
ACCGATAATC	TCGTGATAGT	AAAATTCACC	ATCGTCTAGG	TCATTCAAAT	CTTCCTCAGC	2040
GACCTTGAGA	CTGTATCCCT	TGTACTTTTC	GATAGTATTG	ATATGGTACA	TATCTTTGAA	2100
TTTAATAATG	TCAAAGTTCT	TCTGTTTACG	GTGGCTAGCG	ATGGTCACTG	TTTGGACAAA	2160
CTGATCTTTT	TCATCAAACA	AAACCAGCTC	AGCTCCTTTT	TTAAACCGTT	CTTCTGCAAA	2220
ATCCGTCACA	GACAAGACTC	GCATCTCCCC	CTGTAATCCC	TGCGTATTAA	CGATTTTCCC	2280
AACATTAAAG	TAGTTCATCT	TGTCTCCTGT	AATCTCCTTT	TTCCATCTT	ATTCTAACAA	2340
TTCTCGAATA	ATAGCCGCAA	TTTTTTCCGA	TTCTGACCAT	TGTAAATAAT	GGTGATTCCC	2400
TCCTAAAATG	AGTTTAGTAT	TGGAAGTCCA	ATATTCTGAT	TCTCTGTACT	CTTTTTCTCT	2460
ATAAGGCTGA	CAAAAAACAA	ATACAGGAAT	ATGAGCTTCT	ATAGATACAT	CCTCAAAATC	2520
TTCTCAGTA	ATCTCTCCAG	ATATCTGAAA	TTCTGGATCT	TGATTTTCCA	ACTCTAAGCC	2580
TTTTTCTTGC	ATTAATTCCC	AGATTTTTTT	ATTCGTTTCA	GGACTAAATG	TTGCTTGAGT	2640
TAAGTTCTTA	AAATAAAGTT	CAGGACCACA	CTCGTCAATC	AGCCTCATCT	GCTCTTCCAT	2700



865

TTCTGGATAA GGATTTTCTG AAAAATCAGC AAACATGACT TTTTtagTTG TCGGTTCAAT	2760
TGCTACTAAA GTCTGACGCT TAATTGGTTT CTCGAGTAAT TTGCAAGCTA AAATTCCACT	2820
CCAACTATGT GCACAAAGTA TATATTCAGA AATTCCTAAT TCTTCAAGTA CTTCATAAAC	2880
CGCATCTGCA AGATTATCTA GATTTTTCAGCTTGGTCA TGAATCGGAC TCCTACCTGT	2940
GTTCGGAAAA TCAATTGTCA AATAACCAAT TGTAGGAGGA GGTTTTTCAG GTATAAGTGA	3000
AAAATTTTCA TAACTTGGTA GCAAACCTGC TCCGTTTAAA CAAACTAGCA CTTTCTTTTG	3060
CTTTTGATAA GTAACAGAGA GGCTACCAAT TTCTGTAGAT ACTTCAAACC TCTTCATAAA	3120
GAAATCCACT GATTCTATAT AATGAATTAT TAAAAATCCT TATCCTTTAT TTTATCACGT	3180
TCCAAGGATT TTCTCAAGTT GGAGGAAGGG GACAATATCT CTACTTTCCC TTCAATAATC	3240
CTTCCAAATT ATGTTTATGT TGGTAATTAA TGGCTGCGGT TTTGTCTTTC TCAAAGACAG	3300
TCTTGGTAAG GTCAATATGA TTAATAGCTA CGATTGCGAC GGTGTAGTAA ATGATATCAG	3360
CCAGTTCTCT GGCAAGTTCC TCGTTTGAAT CCTATCCCTT CTTTTCGACC AGAGCGCCTA	3420
TTCAAAACCT CGACTACTTC TCCGACTTCC TCCACTAACT TCATAAAGAG ACCTTCATCA	3480
GTCCGAGACT GCTGTTAATG TTCGATTAAG TAGTCTTGA ATTGCCTAAA CGTTCAATCT	3540
TTTATAGTAT ATTGAAACTA GAATAGTACA CCTTTACTTC TAAAACATTG TTAGAAATCG	3600
ATTTGACTGT CCTGATCGAT TTGTCTGTT CTTGTTTCAT TTTACTATAT CTTCATTCC	3660
ACACAAAAAA GCGAGACATC CGTCCCGCCC TTCTTATTTT TCGTCAATAA CGATTCTTAC	3720
TTTTTTGTAT TCAGTTGGGA CAGAGTAGAC AATCGTTCTT ATCGCAGAAA TACTGCGACC	3780
CTTACGACCG ATTACACGAC CCACATCGCT TTGATCAAGA TTCAAATGAT ATTCCAAAAA	3840
TTCTGGTGTA TCCTCAATCT TGATAGTTAA GGCATCTGGT TGTGAAATTA AGGGTTTCAC	3900
AATCGCAATA ATGAGATTTT CAATCGTATC CATCTGTCAA CCTACTTTAA ACTTATTTTG	3960
AAAATTTAGA ATCGTGGAAT TTTTTCATA CGCCTTCTTT TGAAAGGATG TTACGTACTG	4020
TGTCTGAAGG TTGAGCTCCA TTAGCCAACC ATGCAAGAAC GCGGTCTTCT TTCAAAGTTA	4080
CTTGGTTTTT AGCAACAAGT GGGTTGTAAG TTCCAAGTGT TTCGATGAAA CGTCCGTCAC	4140
GTGGTGAACG TGAATCTGCT ACGTTGATAC GGTAGAAAGG TTTTCTTCTA GAACCCATAC	4200
GAGTCAAACG GATTTTAACT GCCATTTTAA AAGTCTCATT TCTTTAATTT TTTATTTTCG	4260
TGAAATAGCT GAGCTATTTA GCACATGTTT TATTATAGCA GATTCTGGC ATGTGTC	4317

(2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4881 base pairs

866

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

AATTTATTTG ACTGAAAATT GTAGAGGGTT CTCGAAATTT CTTGAATGGT TAAAATAAGG	60
ACAAGAGAAA ACATGGATAT CTATATCCTT GTGCCAAAAA AACCACTGCC CTCCCCAGAC	120
CAACCTGAGG AAAGCAGTGA TTCTTATTTT AGGAGTTAGG AATGAATACA CGAAATCAAT	180
TTAGCTGATT ATTTTTTGTT TTTCAAGAAT TCATCGTATT GTTTTTGCAT TTCGTTCAAT	240
ACTTTTTTCGT AGGCACCTTC AGATTTCAAT TTTTCCATCA ATTCTGGAAT CGCTTTATCT	300
GGGTCTACAG TACCAGTGTT GATAGCTGTA TCAAATGTGT GCATTGTGTT AGCAATAGCT	360
GAGATTTTCAG ATTTACATT GTCAGTATTG AAGATAAATC CAAGCGCTGG AGATTTCTTTA	420
GCTTCTGCCA ATTCTTTCTT AGAATTTTCG ATTTGTGGT CTGTAACGTT TTCGTTGATG	480
TAAAGGATCC AGTTGTTACC AGTGTTCAT CCACCCATGT GAGTGTTCCT TTTGTAGCCA	540
TCAAGAACGC GAACACGGTT TTCTTTACCT TCAATTTTTT CCCAGTTCTT GCCTTCTGGA	600
CCGTAAACAA GACCGTTCAA GAGTTCTGGG TTCGTATTCA AGAGGTTCAA GATTTCCATT	660
GATTTTTCTT TGTTCCTAGA GTTGTGTTGAG ATGACAAAGT TAGCAACTTG TGTGTTTGG	720
TTTTTCTTGA TGAAGTTAGT AATTGGTTTG ATTTGGATAT CTTTGTGGC AACACGTGAA	780
AGCAAGCTGT TACCGTAGTC AGCTGGTCCT ACTGTTTCTT CACGAACGAA CCAAGTATCT	840
TGTTGAAGGT CAAAGGAAGT ATCGCTTCTT GCGACGTCCT TTGGAATGTA GCCAGCTTCA	900
TAGAATTTGT GAAGAGTCTT CAAGTCTTCT TTGAAACGAG GCACTTCGTA ACGGTTTACA	960
ACTTTAGTAG TATCGCCTTC AAGGTCGATA ACGAATGGAA GACCGTTTGC TACTGGGTAG	1020
TCAAAATTAT CAGATGGGAT GAAAACTTTA CCAATAGCAA ATGGTACTAC GTCTGGAGCT	1080
TTTTCTTTGA TTTGTTTCAA GACTGGCTCA AGAGTTTCGT AAGAAGTAAC ACCTGAAATA	1140
TCGATACCAT ATTTAGCAAG GAGAGTTCCG TTGAAGGCAA AGTTTTGAGA TGATGCAACG	1200
TTGGCTGCAA CTGGAACAGC GTAAATCTTA CCATTTACAG TATTACCCTT GATGTAAGCT	1260
GGGTC AAGTG CTTTGTAAG GTCTTTACCT TCTTTTTTGT ACAATTCTGT CAAGTCAGCG	1320
TAAGCACCTT TTTGAGCATT TACAATATAG TTATCTGCAA AGGCAATATC ATAGTTTCA	1380
CCAGATGATG TGATAACTGA CATTTTCTTA CCATAGTCAC CCCAGCCAAG GTATTGGATA	1440
TCCAATTTGG CACCAACTTT TTCTTCAATG ATTTTGTTGG CATTTGCTAA CAATTCATCC	1500
AAGTTGCTG GTTTGTCACC GATTTGGTAC ATTTTGATAA CAGGTTTGTC ACCTGAATCA	1560

867

GCAGCTTTTT TGCTGTTACC TGTCAAATTT CCACAAGCAG CAAGACCTGC AGCCAGAGCG	1620
ACTACACTAG CAGATGCAAA AGCATATTTT TTCCAGTTTT TCATGATAAA AACTCCTTTT	1680
TTTATTTTAA AACTTATAAA CAATGTAATG ATCTTATACT CAATAAAAAT CAAAGAGCAA	1740
ACTAGAAAAC TAGCCGAGG CTGCTCAAAG CACTGCTTTG AGGTTGTAGA TAAGACTGAC	1800
GAAGTCAGTT ACATATATCT ACGGCAAGGC GACGTTGACG CGGTTTGAAT TTGATTTTCG	1860
AAGAGTATTA ACTTCACACA AGGGAAGTTG GGAAGTGAAGA AATGTTATTT CTCAATAAGC	1920
ACTATTCTTT CACACCACCG ATAGTCAAAC CTTTTACAAA GTAGCGTTGG AAAAATGGAT	1980
ACAAAATCGC GATTGGAAGG GTTGCAACCA CAACCATGGC CATACGACCT GTTCTTTTCG	2040
GTAGAGCAAC TCCCAGTTGA CCAATCAAGC CGACCGCTTT GGCAATGTAG TCCATATTTT	2100
GTGGAATTTG CATGAGCAAA TATTGCAATG GATACAAGTT GTCACCTCTG ATGTAAAGAA	2160
GGGCGTTGAA CCAGTCATTC CAGAAACCAA GAGCTGTTAA GAGCGTGATG GTTGGGATAC	2220
CTGGTAGTGA CAATGGCAAA CAGATTTGGA AGAAAATCCG GGCCTCACTG GCACCATCGA	2280
TACGAGCCGA TTCTAGAATG GCTTCTGGAA TGGTCTTCTT GAAGAAGGAA CGCATCAAGA	2340
TGATGTTAAA TGGTGAGAGA AGCATTGGAA CAATCAAGGC CCAAACAGTG TCACCAAGCT	2400
GAAGTACACG GGTCAACCATG ATATAACCTG GTACCAAACC AGCGTTGAAC AACATACTGA	2460
GAAGGACGAA GATGGTAAAG AATCTGCGAT ACTTAAAGGT TGTCCGTGAA ATAGCGTAGG	2520
CATAGGTGCT TGTGATAAAG ACATTTGTCA ATGTCCCAAC TACGGTTACA AAGACAGAGA	2580
TGAAGAGGGC TTGTAGGATT TTATCCTTAA ACTGTGCCAA AAACTCAAAA CCGTCTAAGC	2640
CAAATTGGGA TGGGAAGAAG CTATAGCCGT ATTGGAGGAG GCTTTTCTCG TCTGTCACTG	2700
AAATAATGAT AACGAATACA AAAGGTAGGA TACAAGAGAG GGCAATCAAA CCCGAAATGA	2760
TACTGAAGAA GATATCTGCT TTCTTACTGA AGGAGTGAAT GCCGACATTA TCAATTTTTT	2820
CTTTTTTAAT TTTCTTTTTT GCCATATTCT CCTCCTTTCT AGAACAAAGC TGAGTTTGGA	2880
TCGACTCGTC TTGCAAGCAA GTTTGATAGG ATAACCAGAA TCAAACCAAC AACGGATTGG	2940
TAAAGACCGG CTGCTGCAGC CATACCGATA TCTGCTGTCT GAGTCAAACC ATTAAAGACA	3000
TATACGTCCA AAACGTTGGT TACATTGTAA AGCTGACCAG CATTGTGTGG GATTGTATAG	3060
AAGAGACCGA AGTCTGCGCG GAAGATATTT CCGACTGCAA GGATGGTCAA TACAGTTACA	3120
AGCGGAGTCA ACTGAGGAAT GGTACGTTG CGAATACGTT GCCACTTGCT AGCTCCGTCC	3180
ACTGTGCGTG CTTGCTAGTA GGTGGATCA ATTCCCATGA TCGTGCATA GTACATGACA	3240
CTGCTATATC CAAAGCCTTT CCAAATACCT AGGAAAAGTA GGAGATAGGG CCAGATGCCC	3300

868

AGGTCAGCGT	AGAAAATTGAC	TTCTTTGAGA	CCAAGACTTT	CCAATAGATG	ATTGAACACC	3360
CCTTTATCAA	TATTTAGGAA	GGCATCTGTA	AAGAACTGA	TGATAACCCA	AGACAAGAAG	3420
TAAGGGAACA	ACATAGAAGT	TTGAAAAATC	TTCAACCATC	TCTTAGAACG	GAGCTCGCTG	3480
AGGATAATGG	CAATCCCTAC	AGATACAACT	AAACCTAGAA	AGATAAAGCC	AAGATTGTAG	3540
AGGACAGTAT	TTCGTGTGAT	AATAAAGGCG	TCTCTTGAAC	TAAATAAGAA	TCTAAAATTA	3600
TCGAGTCCGA	CCCATTTACT	ATTTATGATA	CTATCTATGA	AACCATTACT	GGTCATGTGG	3660
TAGTCTTTGA	AGGCAACCAC	GTTCCCAAAT	ACTGGAATGT	AAAAGAATAG	AATCAACCAG	3720
AGTGCCCCCTG	GCAAAACCAT	CAAGAGAAAG	ATCCAGTTGT	CTCTCAATGT	TTTTGAAAAC	3780
TTTTTCATAA	TTTCCTCCCT	TTTTATTTTG	ATATCCATCT	AAAAATTCTT	TTTTAGACTT	3840
TTGATAACGA	TTACATTATT	AGTATACTCC	TATTTGCAGG	TTAGGTAA	CTCCTAATTA	3900
TAGAAAAAAC	TCCACAAATT	ATGTAGCAGA	TTTAAACTT	TATCACCCT	ATCAACAAA	3960
TGTCCTAAAT	CAATTGTTTA	TTTTATCTCT	ATTAGCCCAG	TGATGGCGTC	ACTCTGTTAT	4020
AAGCATCCAA	CAACGGGGTA	TACTGAAAAA	TCTCCAGACT	AGGGAACCTA	GCGATAGTTC	4080
CTAATCTGGA	GATTTTTAAT	ATGTTATTAG	GCGTTTGCTT	TCAACTTAGC	AATAACCTCT	4140
TTAAGATTAT	CAATCACTC	TGCTGCAGTA	TGCTCAGAGC	CTTTTTCATC	TGCCAAGAAC	4200
AAAACTGCTT	TTTGAAGTTC	TTTTTGAGAG	TTTTCAAGGA	CATCCTTATC	TACTGTTTCA	4260
AGGTTTGACT	CTTTAAGAAG	TTTACTTAAT	TCCTTGCGTA	ATTTCTTGAG	TTTGATTTGC	4320
AGACTCATCT	TCTCCTGCTG	TTTCTTTGCC	CGCTGTTTGT	CCTCCATCCT	TAGTTGCTGA	4380
CTGGCTTTCC	TTAATGGACT	CTAGGGAAGC	AATGGCATCT	TTGACTGTTT	GCAAGATATC	4440
ACGTAAACCT	TGCTCTGTCA	AACTATCATC	TGCAAAAGCT	TTATTAGCCT	CTGCCAAAAC	4500
CAGACGTGCT	GAATCTGTGG	TAGGATTCTGA	TACACCTGTC	AATGATCTCA	AAAGATTTTC	4560
TAAGGTTTGA	GTCTGCTTAC	TAATACTAGA	CTAAATCAA	AAAGTATTAT	ATAACAGTGA	4620
TATGAAATCA	ACTAAAGAAG	AAATCCAAAC	CATCAAAACA	CTTTTAAAAG	ACTCTCGTAC	4680
AGCTAAATAT	CATAAACGCC	TTCAAATCGT	TCTATTTTGT	CTGATGGGCA	AATCTTATAA	4740
AGAGATTATA	GAACTTTTAT	AGTAGTTTGA	AATAAGATGT	GAACATCTCT	ATCAGGAAAG	4800
TCAAATTAAT	TTATAGAAAT	ATTTTAGCAG	CCAAGGTGTA	CTGTTATAGA	TTCAATACAC	4860
TATACTTGGT	GGTTTAGCTC	G				4881

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11121 base pairs

(B) TYPE: nucleic acid

869

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

AGGATCCCCG GAAAAGGAGA CTAAAAATGA AGAAAAAATT TCTAGCATTT TTGCTAATTT	60
TATTCCCAAT TTTCTCATTG GGTATTGCCA AAGCAGAAAC GATTAAGATT GTTCTGATA	120
CCGCCTATGC ACCTTTTGAG TTTAAAGATT CAGATCAAAC TTATAAAGGA ATTGATGTTG	180
ACATTATTAA CAAAGTCGCT GAGATTAAAG GCTGGAACAT TCAGATGTCC TATCCTGGAT	240
TTGACGCAGC AGTCAATGCG GTTCAAGCTG GGCAAGCCGA CGCTATCATG GCAGGGATGA	300
CAAAGACTAA AGAACGTGAA AAAGTCTTCA CCATGTCTGA TACTTACTAT GATACAAAAG	360
TTGTCATTGC TACTACAAAG TCACACAAAA TTAGCAAGTA CGACCAATTA ACTGGCAAAA	420
CCGTTGGTGT TAAAAACGGA ACTGCCGCTC AACGTTTCCT TGAAACAATC AAAGATAAAT	480
ACGGCTTTAC TATTAAAACA TTGACACTG GTGATTTAAT GAACAACAGC TTGAGTGCTG	540
GTGCCATCGA TGCCATGATG GATGACAAAC CTGTTATCGA ATATGCCATT AACCAAGGTC	600
AAGACCTCCA TATTGAAATG GATGGTGAAG CTGTAGGAAG TTTTGCTTTC GGTGTGAAAA	660
AAGGAAGTAA ATACGAGCAC CTGGTTACTG AATTTAACCA AGCCTTGTCT GAAATGAAAA	720
AAGATGGTAG TCTTGATAAA ATTATCAAGA AATGGACTGC TTCATCATCT TCAGCAGTGC	780
CAACTACAAC TACTCTCGCA GGATTAAAG CTATTCCTGT TAAGGCTAAA TATATCATTG	840
CCAGCGATTG TTCTTTTGCC CCTTTTGTTT TCCAAAATTC AAGCAACCA TACTCTGGTA	900
TTGATATGGA ATTGATTAAG GCAATCGCTA AAGACCAAGG TTTTGAAATT GAAATCACCA	960
ACCCTGGTTT TGATGCTGCT ATCAGTGCTG TCCAAGCTGG TCAAGCCGAT GGTATCATCG	1020
CTGGTATGTC TGTCACAGAT GCTCGTAAGG CAACTTTTGA CTTCTCAGAA TCATACTACA	1080
CTGCTAATAC CATTCTTGGT GTCAAAGAAT CAAGCAATAT TGCTTCTTAT GAAGATCTAA	1140
AAGGAAAGAC AGTCGGTGTT AAAAACGGAA CTGCTTCTCA AACCTTCCTA ACAGAAAATC	1200
AAAGCAAATA CGGCTACAAA ATCAAAACCT TTGCTGATGG TTCTTCAATG TATGACAGTT	1260
TAAACACTGG TGCCATTGAT GCCGTTATGG ATGATGAACC TGTCTCAAA TATCTATCA	1320
GCCAAGGTCA AAAATTGAAA ACTCCAATCT CTGGAAGTCC AATCGGTGAA ACAGCCTTTG	1380
CCGTAAAAAA AGGAGCAAAT CCAGAACTGA TTGAAATGTT CAACAACGGA CTGCAAAACC	1440
TTAAAGCAAA CGGTGAATTC CAAAAGATTC TTGACAAATA CCTAGCTAGC GAATCTTCAA	1500
CTGCTTCAAC AAGTACTGTT GACGAAACAA CGCTCTGGGG CTGCTTCAA AACAACTACA	1560

870

AACAATCCT	TAGCGGTCTT	GGTATCACTC	TTGCTCTAGC	TCTTATCTCA	TTTGCTATTG	1620
CCATTGTCAT	CGGAATTATC	TTCCGGTATGT	TTAGCGTTAG	CCCATACAAA	TCTCTTCGCG	1680
TCATCTCTGA	GATTTTCGTT	GACGTTATTC	GTGGTATTCC	ATTGATGATT	CTTGACAGCCT	1740
TCATCTTCTG	GGGAATTCCA	AACTTCATCG	AGTCTATCAC	AGGCCAACAA	AGCCCAATTA	1800
ACGACTTTGT	AGCTGGAACC	ATTGCCCTCT	CACTCAATGC	GGCTGCTTAT	ATCGCTGAAA	1860
TCGTTCTGG	TGGTATTCAG	GCCGTTCCAG	TTGGCCAAAT	GGAAGCCAGC	CGAAGCTTGG	1920
GTATCTCTTA	TGAAAAAACC	ATGCGTAAGA	TTATCTTGCC	ACAAGCAACT	AAATTGATGT	1980
TGCCAAACTT	TGTCAACCAA	TTCGTTATCG	CTCTTAAAGA	TACAACTATC	GTATCTGCTA	2040
TCGGTTTGGT	TGAACTCTTC	CAAACTGGTA	AGATTATCAT	TGCTCGTAAC	TACCAAAGTT	2100
TCAAGATGTA	TGCAATCCTT	GCTATCTTCT	ATCTTGTAAT	TATCACAATT	TTGACTAGAC	2160
TAGCGAAACG	CTTAGAAAAG	AGGATTCTGT	AATGGCAAAA	TTAAAAATTG	ATGTAAATGA	2220
TTTACACAAG	CACTATGGAA	AAAATGAAGT	CCTAAAAGGA	ATTACGACTA	AGTTCTATGA	2280
AGGAGATGTT	GTTTGTATCA	TCGGTCTTTC	AGGTTCTGGT	AAGTCAACTT	TCCTCCGTAG	2340
CCTCAATCTT	TTAGAAGAAG	TCACTAGCGG	TCACATCACT	GTGAACGGCT	ATGATTTAAC	2400
TGAAAAAACA	ACCAATGTTG	ACCACGTCCG	TGAAAATATC	GGCATGGTAT	TCCAACACTT	2460
CAACCTCTTC	CCTCATATGT	CTGTATTGGA	CAACATCACC	TTTGCTCCTA	TTGAGCACAA	2520
GTTGATGACT	AAGGAAGAAG	CTGAGGAATT	GGGAATGCAG	TTGCTTGAAA	AGGTTGGACT	2580
AGCAGATAAA	GCTAATGCCA	ATCCAGATAG	CCTATCAGGT	GGTCAAAAAC	AACGTGTGGC	2640
CATCGCTCGT	GGCCTAGCAA	TGAATCCAGA	CATCATGCTC	TTCGATGAAC	CAACTTCTGC	2700
CCTTGACCCT	GAGATGGTTG	GAGACGTACT	TAACGTTATG	AAGGAATTGG	CTGAGCAAGG	2760
CATGACCATG	ATTATCGTAA	CCCATGAGAT	GGGATTTGCT	CGTCAGGTTG	CCAACCGCGT	2820
TATCTTTACT	GCAGATGGCG	AGTTCCTTGA	AGACGGAACA	CCTGACCAAA	TCTTTGATAA	2880
CCCACAACAC	CCTCGTCTGA	AAGAGTTCTT	AGATAAGGTC	TTAAACGTCT	AAACTCAAAC	2940
TGTAAGGATT	TCCTTGCACT	TTTCTACCT	CGTATTGGAA	TTTTTGATTT	TTCGGAAAAAT	3000
TATGTTAGAA	TTAAGTTTAT	GAAATGAGGT	TTCTCATAC	CTAGCAAGAC	TAGGAATAAA	3060
AATAGAAATT	AGGTAGCTAG	ATGTCATCTA	AGGTTATTGT	TACAATTTTC	GGTGCAGGTG	3120
GAGACCTGGC	TAAACGCAAG	CTCTACCCTT	CCCTTTTATG	ACTATATCAA	TCCGGCAATC	3180
TTTCCAAGCA	CTTTGCCGTT	ATTGGAACGT	CCCGTAGACC	TTGGAGTAAG	GAATATTTTG	3240
AATCTGTAGT	TGTCGAGTCC	ATCCTTGATT	TGGCAGATAG	TACCGAGCAA	GCCCAAGAAT	3300
TTGCTAGCCA	CTTCTACTAT	CAAAGCCATG	ATGTCAATGA	TTCGGAACAT	TATATTGCTT	3360

871

TGCGTCAATT ACAAGCTGAG CTTAATGAAA AATACCAAGC TGAACACAAT AAGCTCTTCT	3420
TCTTGTCTAT GGCACCTCAG TTCTTTGGAA CCATTGCCAA ACACCTCAAA TCTGAAAACA	3480
TTGTCGATGG CAAAGGTTTT GAGCGCTTGA TCGTTGAAAA ACCATTTGGT ACAGATTACG	3540
CAACTGCAAG CAAGTTGAAT GACGAACTCC TAGCAACATT TGACGAAGAA CAAATTTTCC	3600
GTATCGACCA TTATCTTGGT AAGGAAATGA TCCAAAGCAT CTTTGCAATT CGCTTTGCAA	3660
ACTTGATTTT TGAAAACGTT TGGAAACAAGG ATTTTATCGA CAATGTTCAA ATTACCTTTG	3720
CGGAGCGCTT GGGTGTAGAA GAACGTGGTG GCTACTATGA CCAATCCGGT GCCCTCCGTG	3780
ACATGGTCCA AAACCACACT CTACAACCTC TTTCGCTCCT CGCCATGGAC AAACCAGCAA	3840
GCTTCACAAA AGACGAGATT CGTGCTGAAA AGATTAAGGT CTTTAAAAAC CTCTATCATC	3900
CAACTGATGA AGAACTCAAA GAACACTTTA TCCGTGGGCA ATACCGCTCT GGTAAGATTG	3960
ATGGCATGAA ATACATCTCT TATCGTAGCG AGCCAAATGT GAATCCAGAA TCAACAACTG	4020
AAACCTTTAC ATCTGGTGCC TTCTTTGTAG ACAGCGATCG ATTCCGTGGT GTTCCTTTCT	4080
TTTTCCGTAC AGGTAAACGA CTGACTGAAA AAGGAACTCA TGTCAACATC GTCTTTAAAC	4140
AAATGGATTC TATCTTTGGA GAACCACTTG CTCCAAATAT TTGACCATC TATATTCAAC	4200
CAACAGAAGG CTTCTCTCTT AGCCTAAATG GGAAGCAAGT AGGAGAAGAA TTAACTTTGG	4260
CTCCTAACTC ACTTGATTAC CGTACAGATG CGACTGCAAC TGGTGCTTCT CCAGAACCAT	4320
ACGAAAAATT GATTTATGAT GTCCTAAATA ACAACTCAAC TAACTTTAGC CACTGGGATG	4380
AAGTTTGTGC GTCATGGAAG TTGATTGACC GTATTGAAAA GCTCTGGGCT GAAAAAGGTG	4440
CCCCACTTCA TGAATAAAA GCTGGAAGCA TGGGACCTCA AGCCAGCTTT GACCTACTTG	4500
AAAAATTCGG TGCCAAATGG ACTTGGCAAC CAGATATCAC CTATCGTCAA GATGGTCGCT	4560
TAGAATAAAA AAATTTCTTG CAAGTTTATG CcTTGCAGGA TTTTGTCTTC TGATTAGATT	4620
AAACCTTCCA AGAGACCTTT CATAAAGTTT TCTGAGTTAA ACTCTCCAAT ATCATCGATT	4680
TTTTCAACAA AACCAATCAA TTTTACAGGA ATATTGAGTT CTTACGAAT GGCTAGAACC	4740
ACACCTCCTC GAGCAGTTCC ATCAATCTTA GTCAAAACAA TTCCCGTTAA AGGTGTGATT	4800
TTCGAAAAAT CTTTGGCCTG TACTAGGGCA TTTTGACCTG TTGATGCATC AAGTGCCAAG	4860
AAGGTTTCAT GTGGTGCTTC TGGCACAACA CGTTTGATAA TACGACCAAT CTTTCCAAAC	4920
TCAGCCATAA GGTATCCTT ATTTTGAGCA CGACCAGCAG TATCAATCAT GAGAATATCG	4980
ATACCTTCAG TCACGGCAGG TTCCATACCA TCAAAGACCA CGCTGGCTGG ATCAGCTTTT	5040
TCAGGTCCAG TTAATACTGG AACATCTACT CGTCGGCCCC ATTCAGCTAG CTGAGCTACT	5100

872

GCACCCGCAC GGAAGGTATC TGCTGCAACC AGCATGACCT TCTTACCAGC TTGTTTGTAG	5160
CGGTGGGCTA GTTTTCCGAT AGAAGTTGTT TTCCCAACAC CATTACACACC AACAAAGAGC	5220
ATAACTGTCA AGTTATCTTG GAAGTGGATG CTTTCATCGT AGCTACCATC CTTTTCATAA	5280
AGCTCAACCA ATTTCTCAAT GATGACACGA CGAAGTACAT CAGGTTTCTT GGCATTTTCA	5340
AGCTTGGCTT CGTAACGTAG TTCCTCCGTT AAGTTAGAAG CGACTTGGAC ACCAACATCA	5400
CTCATAATCA GCAGTTCTTC CAGTTCCTCG AAAAATTCTT CGTCAACAGA GCGGAAGTTA	5460
GCAAAGAAGG CATTCAAGCG GGCACCGAAA CCTGTGCGAG TTTTCTTAAG ACTGCGGTCA	5520
TATTTTTCCT GAACAGTTTC TTCTGTTTGA GGAGCTTCTG GTTCAAGCAC TTCAGAATTA	5580
TTTTCTTCTA CAGTTCCTTC GTGCTCAAGC TTCTCTTCCT CTGGTAATTC TTCTGAGTTT	5640
GGTAATTCTT CTATTTCTTC TTGAGAAACC CCTACAGCTG GCTCTGAATC CTGACTTTCT	5700
TCAACTGTGT CTTGGATTTC CTCTTCTTGG AACACAGCTT GTTCAACAAT TTCAACCTCT	5760
GCTTCTTCCT GAGAACTTC CTCAACTTCT GTGAAGGTAG GATCAACATC TTCAGACAAA	5820
TCAAGATTTT CCAGAGCTTC TTTTACAAC TCTTCGATTT TAGGTTCTTC TTTTTTCCG	5880
AATAGACGGT CAAACAATCC CATATCTTAG TTCTCCTTTA GCACATATTC TTCGATAGCC	5940
CAGGCGACAG CTTCTCATC GTTGGTCATC GCGGTCATA CATTGCGGC TGCCTTTACT	6000
TCAGGAACAG CGTTTTGCAT AGCAACACCA AGACCTGCCC ATTCAATCAT AGAGAGGTCA	6060
TTGGCCTCGT CACCACAAGC CATCACTTGA CTTTGGTCGA TTCCAAGATG GCTGATTAGT	6120
TTTGCCAAAC CTGTTGCTTT ATGAACATTC TTTGGTGACC ATTCTAGCAA CATTTACGTT	6180
GATTTAAAGA TTTCATATTG GTCAAACAAT TCTGGAGAAA TCTTCTGAAT GGCTGCATCC	6240
AAGGGTTCTT GAGCAAAGGC AGTCACGCAT TTGTTGTAGG TCATTTGACT AGATAAGTCT	6300
TCAAAGTCCA CTGGAACAAA GGTCAAAGCT GGATTGAATT TGGCATAAAG ACTTTCTTGG	6360
TCCGATTGGA TTTGATAAAC TGTTCTTCTT GAGATGGCAT CAAGAGGCAG TGATAATTTT	6420
TCTGTTTCTT CATACAAACG TGCCACATCA TCATATGAAA AGACTGTTTT ATCAAGGATT	6480
TCTCCTGTAT TTTTCTGAAC TAATCCACCA TTAAGGTAA TGGTATACTC ATCTTCCTGA	6540
CCGTCAGTCC CTAATCATG GAGAAAGAAA TCCATGGCTT TTAAGGGACG ACCAGTTGTC	6600
AATACGACCT TGATACCAGC ATCACGCGCA GCTTGAAGG TTTCTTGGT ACGATCCGTC	6660
AGCCTTTTAT CAGTAGTCAG CAAGGTCCCG TCCAAGTCCA ATGCAATCAA TTTTATATCT	6720
GCCATTATAA GCCCTCCATA TAAGCTATAA CCGACCGTTC CTTATGGTGA CCAATCACAG	6780
TCTTTGCTAA TTCTAAAATT TCAGGTCGTG CATTTTCAGG AGCTACAGGA TGTCCACAA	6840
CCTGCATCAT ATGTAAGTCA TTAAGATTGT CTCCTAAAAGC CATGACCTGA TCCATTGTGA	6900



873

TACCAAGTTT TTAACTAAT TCAACAATGG CCACTCCCTT ATCGACATAG TCCAGAACAA	6960
TATCAATGGA TTCAAAGCCA GTTGTCATGG CCTTAACACC AGGAACGTTT TCGTTTACCC	7020
AAGCCTCCCC ATCTTCCAGC GTTCTTCTG TGAAGTTGGT TGTAAATTTG AAAATGTCAT	7080
CTGTGATATC TTCCAACTC GCTACTTTTT GGATATTTTC ATTATAGTGC TGACTCACTT	7140
TCAAATAGGT CTCATCAACC GTATCTAGAA CATATGAACC CTTCTTACCC GTCAAGAGCA	7200
-----GTTTATTGAT ATCTACATAA GGTGAAGTTT TCAGCTTTTC AAAAGTTGCC AGATAAAAAGT	7260
CACGAGACAT AGTCGCTTCA TACAAGTCCT GACCTTGATA CTCTACCAA CTGCCATTTT	7320
CCGCGATGAA AATAATGTCA TCACGAACAC CAGCAAATAA TTTTCTAGA GACAGAAATC	7380
CCCGACCCGA AGCTACCGCA AAGTAAATCC CTTTTCTCTT GTAGGAAACC AAGAGAGACT	7440
TGAGACGATC CATATCAAAG CGTCCATTCC CATCTAGGAA GGTTCCTGCC ATATCCGTTG	7500
CTACTAGTTT AATTGTCATC CTTCAATACT TTCTAAATCT TTAACTTAA CTGAAACAAT	7560
CTTTGAAACA CCCGATTCTT GCATGGTCAC TCCATAGATG GAATCAGCCG CTGCCATGGT	7620
TCCCTTACGG TGGGTTACGA CGATGAACTG GCTGTCCTTG TCAAAGCGGT TGAGGTAATC	7680
CCCCAAACGT TTAACATTGG CTTTCATCCAG CGCAGCTTCC ACCTCATCCA AGATAACAA	7740
TGGAATAGTC TTGACACGAA TAATGGAGAA GAGCAAGGCA AGAGCCGATA GGGCTTTTTT	7800
ACCACCACTC ATGAGATTAA GAGACTGGAT TTTCTTGCCT GGTGGTTGGA CAGAAATTTT	7860
AACCCAGCT GTCAGCAAGT CTCCTTCAGT CAAAATGAGG TCAGCCTGAC CTCCACCAA	7920
CATCTGCTTG AAGGTCATT TAAAGGACTC ACGAATGACC TCAAAGGTTG ATTTAAAGCG	7980
TTCTTGACC TCATCATTC TCTCTGTAAT GGTCTCAAGG AGCAGGTTTT TCGCAGACAA	8040
AATATCATCA CGTTGGCTAT TTAGGAAATC CAGACGGTTG TGAATTTCTT CGTACTGTTT	8100
AATAGCGTCT AAATTGACAG GACCCAGTGA GCGTATAGCC TTCTCTAAAT CCTTAACTTC	8160
TTGCTCTGCC AGATTGAGAT TTTCCAACTC ATGCGCCTTT TCTAAAGCTT CTGTGTAGCT	8220
GATCTGGTAC TGGTCTGTTA ATTGACTTTG TAGATGGCGC AAGCGCTCGC TAACCTTTTC	8280
TTTCTTGGCT TCAGCACGAG TTTGCTTGGC AATCCACTCT TCATTCTGCT GGCGAGCCTG	8340
ATCCAAATGA CTAGCAATAT CATCCAGTTG ACCCTCAATA TCATCCAAT CAACTGCTT	8400
GCGAATCAA CCTTGTGGA GATTTGTTTT TTGAGTTTTG GATTCTTCCG CCTGTTGACT	8460
GAGCAATTCT GTATCAACCT TCTCAAGATT ATCAATCTTT TCTTGAAGAA GCGCTGGAT	8520
TTCTCTTGT TCAAAATCAA GATTGTCCAA TTCCTTGCTT AAGCGTTCAA TATCAGCAAC	8580
TTCATAACGT TTTTGCCCTT GCAGTTCTGT CTTAAGCAAA CGAGCTTGCG CTAGCTCTTC	8640

874

CTGCAAGTTT TGATAGCGTT CTTGGATGGC ATTTTGTGTA GACTTAATCT CTTCAATCTC	8700
AGCTTCCAGA TTTTGCTTGT CACTGGAGAT TGCAGCAAGA CGCTCTTGGC AGTTTTCCTT	8750
ATCCGCTTGC CAATCTCCCT CGGAAAGACG ATCTATTTCC TCTTCTTGGG GTTTCCAAAG	8820
AGTTTCCAGT TCTTCAACTT GCTGACTAGT TTGCTGATAA GCGAGGAACA AGCCTTGCTC	8880
CTGAATACGT GCCTGCTCTC CTTGAGATTT AATAGCTTCT AATGACTCGG TCAATCTGGC	8940
CATCTCATCT TGCAAGGTCT TCAAAGTCGC CTCTTCTGAA CCCAAGCTTG CTTCTTCTTC	9000
AGCAATTTCT TTTTGTAATT GCTCCAGTTC TGGCTTGATA AAAATGCTGT TATTCTGGCG	9060
ATTGGCACCA CCTGCATAAG AACCACCTGT GCGCAACTCT GTCCCATCCA ATGTCACCAT	9120
ACGAACCTGA TAACGAACTT GCGGAGCTGC TGCACGCGCA TGTCTACGG TATCAAAGAT	9180
AGCCGTCGTA GCTAGCAAGT TCTTGAAAAT GGCTTCCAGT CTAGTATCAA AAGTCACCAA	9240
CTCATCTGCC ATCCCAAGGA AACCTGGGCT TACAGCGATA GCATCTTGGT TCTGACTAGA	9300
AATCGTACGC GCCTTGATAG TGGTCAAAGG AAGAAAGGTT GCACGACCGG CTCTGTTCCG	9360
TTTAAGGAAG TCAATAGCCT TGGTTGCCGA CTCTTCATCT TCTACGATGA TATGCTGGCT	9420
ACTTGCCCCCT AAGGCAATCT CTAGGGCAGT TTGATAATAA ACATCAAAGG TCAGATGCTC	9480
ACTGACTGCA CCAATAATCC CACCTAGGCG ATCTTTTTCT TGGAGAACAC TCTTAACACC	9540
TGCATAAAAG TTACTATGAT TTCTCAGGAT ATTTTCCAAA CTTTGAGCTC TGGCCTGCTT	9600
GTTTTTGAGA TTATCCAGAC GGTCAAAGAG TTGGCTTTGT TGAGCTTGAT AGGAAGTTT	9660
CTGCTCCTCT TGCTCCTTGG CAATAGCTTG GTAGTCAGCC AATAATTTCT GAACCTGCTC	9720
CTTGGCACTT TCAAGCTCTT CTTTTGCTG ACTAGCCTTC TCTTTAGCTA TAGCTAATTG	9780
CTCTTTCAGC TTTTCTAGTT GATCTGCTTG TTTTGTAGAA AGCTGACGAC TATTTTCCAA	9840
CTCATTCTCA ATACGGGTCA ACTGGTTTGA GACATCCGCT TCTTCTTGTA AAAGAGCTAC	9900
AAAGCGTTCA CGTAAGAGCT CAATCATCTG ATCAGGATCG TCTGAGAAAG CCAGCAATTC	9960
AGCTTCTAAA CGATTGAGTT TTTGATTATT TTGGACTAGA TTTCCCTCTA ACAGAGCTAA	10020
AGAGCTTTCT TTATCAGACT TTTCTTTGCT GAGTGAATTT CTCTTATCCT CCAAAGCAGC	10080
CAAACGGGCT TGTGCCTCCT GTTGATTCAA GGCCACTTGC TCGGACTCCA GTTTCGATAG	10140
GGCTAATTTT CTTTCTAAAT CACTAATCAG ACTAGTCAAG TCCATCAAAC TGCCTTGGTC	10200
TTTGGCCATT TCAGCCTGTA AATCTTGGCG TTGCTTTTGA AGAGTTTGAT TTTCTTCTTC	10260
TAATTTTCA CGCTTTTGGT AATAACTCAT CAAGAGTTCT TGAACCTGAG TCAACTCTTC	10320
TTCTGTCGAC TCTAGTTCAG CCTTATTTTC CTGATTTGA GCAACCAGAA CATCTAAATA	10380
AATAGCCTTA CGTTGTCCTT CCAAGTCTAA AAACCTACGG GCATTCTCAG CTTGCTTCTC	10440

875

AAGAGGCTTG	ATTGATTAT	CCAACTCGTA	GATAATGTCC	TCTAAGCGGT	CCAGATTATC	10500
CTGAGTTTGC	TGCAGTTTAC	TCTCGGTTTC	TTTTCTGCGA	GTCTTGTATT	TTAAAACTCC	10560
AGCAGCTTCT	TCAAAAATAG	CTCGTCGTTC	CTCAGGCTTG	GAATTAAAAA	TCTCCTCAAC	10620
CTTCCCTTGG	GAAATAATAG	AGAAGGAATC	TCGTCCCAAT	CCAGTATCCA	AGAAGAGGTC	10680
ATGAATATCA	CGCAGACGGA	CTTTCTTGCC	GTCAATCTTG	TATTCGCTAT	CTCCACTACG	10740
ATAGACATGG	CGTTCACCCC	TGATTTCTTG	ACCTGCATCC	TTGATAAATC	CGTCATGATT	10800
ATCCAGAGTC	ACAACTACAG	AAGCATAATT	GAGCGGTTTG	CGACTTTTCGG	TTCCAGCAAA	10860
GATGATATCC	GGCATCTTGC	CCCCACGGAG	ACTCTTGACA	CTAGACTCCC	CCAAAGCCCCA	10920
ACGCAGACTT	TCTGTAATAT	TGGACTTTCC	AGATCCATTG	GGTCCAACAA	CTGCCGTCAC	10980
ACCTTGGTCA	AAAACGACCT	TGGTCTTATC	AGCAAAAGAC	TTGAACCCCT	GAATTTTCGAT	11040
TTCTTTTAAA	TACATGAATC	CAGCCCCTTC	TCAACGGCAT	TTTTGGCAGC	TTCTTGCTCT	11100
GCTAATTTCT	TAGAACGACC	TTGGCCTTGA	CCGATGCTCT	TACCTTCAAC	AAGAACTTCT	11160
ACATCAAAAA	CCTTATCGTG	AGCAGGCCCT	GTTTCAGAAA	TCACCTGATA	ACGAATAGCC	11220
ACATCACCAT	TGACCTGAAG	CAACTCTTGG	AGATGGGTTT	TATAGTCTGT	AATCATCTCA	11280
AACTCGCCTG	CTTCAACCTT	AGGAATCATG	ACTTGATAGA	TAAATTCCTT	GACCTTGGCC	11340
ACATCCTTAT	CCAAAAGAAG	GGCACCAAGA	AAGGCTTCAA	AGGCATCACC	AAGAATGGTG	11400
TCACGATTGC	GACCACCTGA	TTTTTCTTCC	CCTTTACCCA	ACTTGATAAA	CTGGTCAAAC	11460
TGGCAATCAC	GCGCAAAACC	AGCTAAACTC	TCCTCACGGA	CAATCATAGC	ACGGAGTTTT	11520
GATAGGTCAC	CTTCAGGCTT	TTTAGGATAT	TTTTTATATA	GATATTCTGA	AATCAATAAC	11580
TGTAGAACAG	CGTCTCCTAA	AAATTCCAAG	CGTTCATTGT	GTGAAATTTT	TAAGAGGCGG	11640
TGCTCATTTG	CATAACTCGT	ATGAGTAAAG	GCAGTTTCCA	GTAACCTTTT	GTCTGCAAAT	11700
TCGATTGCAA	AATGATTCTT	TAGTACAGTT	TGTAATTCTT	TCATACCAAC	CTCTTTCTAA	11760
CTGATAATAG	TCCTTTTTAT	TATATCAAAA	AAAGCCCCCT	GAGTCACTCT	AAAACGGGAC	11820
TGGAAAGCAT	TTGGGAATTC	TTTAGACAGA	GATTCTCAGT	TTTAGCGGCA	AATTTGGGTC	11880
AGGATAAAGA	AAAAAGCCCT	ATTAAAGGCT	TTTTAGGATG	TTTACATCCA	CCCTGAGGGA	11940
ATCGAACCCC	CATCTCAAGA	ACCGGAATCT	TACGTGATAT	CCATTACACT	AAGGGTGGAA	12000
ACTTGTTTTA	TTATAACAGA	AATTTGCTCT	AATAACAAGT	TTTTTGGTCA	AAGACCCCGT	12060
CTTAGTGGGA	AGCATCCCCA	TTCCAGATGG	AGTTTTTCAC	GATCACATAA	TCAACGTGTT	12120
TAAGGTCAGC	AACCTGACGT	CCACCTGCAT	AAGAAATAGC	ACTTTGAAGG	TCTTGTTCCT	12180

876

TCTCAGTTAA AGTGTCTTGC AGATGACCTT TAGCAGGAAG CAAGATACGT TTGCCTTCCA	12240
CATTTTGTGA AGCACCTTTT TGATATTGTG AGGCTGAACC ATAATATTCT TTGAACTGTT	12300
CACCATCGAC TTCAATCGTT TTCCCTGGAC TTTCAATGTG TCCTGCAAAG AGGGAACCAA	12360
TCATGATCAT GCTAGCACCG AAGCGGATAG ACTTAGCAAT ATCACCGTGA GTACGAATTC	12420
CTCCATCAGC GATAATCGGT TTACGCGCAG CCTTGGCACA CCAGCGTAGA GCAGCCAACT	12480
GCCAACCACC TGTACCAAAA CCAGTCTTAA CCTTGGTGAT ACAAACCTTA CCAGGACCGA	12540
TTCCGACCTT AGTAGCATCC GCACCAGCAT TTCCAATTC ACGCACAGCT TCTGGTGTTC	12600
CCACATTTCC AGCAATGACA AAGGTATCTG GCAATTCTTT CTGATGTGT TGAATCATAG	12660
AAATCACGCT ATCCGCATGA CCATGAGCAA TATCAATAGT GATATACTCA GGAGTATCAG	12720
CCTTGAGCTG GCTAACAAAA TCATACTCAT AATCCTTAAC ACCGACAGAG ATAGAAGCAA	12780
TGAGCCCTTG ATTGTGCATT CGTTTAATAA AAGGAATGCG TCCTGCCTCA TCAAAACGGT	12840
GCATAATGTA GAAGTAACCA CCTTTAGCCA GTTGCTCTGC TACATTTTCA TCCAAAATCG	12900
TCTGCATATT CGCTGGCACA ACAGGTAGTT TAAAGGTGTG ATTTCTTAAA GTGACACTTG	12960
TATCCGCTTC TGCACGGCTT TTAATGACAC ATTTATTGG AATCAATTGA ATATCTTCGT	13020
AATCAAAAAT TGGAAATTC TTTAACATAT CGATGTCTCG TTTCTTTTGT AATGACCTAC	13080
CTATGCTCTT GCATCACTAC GCCTTTTCCG ACGTTTCCTG G	13121

(2) INFORMATION FOR SEQ ID NO: 127:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9578 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

CCGAATGCAA TGTTTACGGT TGAAGTTGAA AATGGACATC AGATTTTAGC AACAGTTTCT	60
GGTAAAATTC GTAAAACTA TATTCGTATT TTAGCGGGAG ATCGTGTTAC TGTCGAAATG	120
AGTCCATATG ACTTGACACG TGGACGTATC ACTTACCGCT TTAAATAATC GAAAAACTTG	180
GAGGGATAAG AAATGAAAGT AAGACCATCG GTCAAACCAA TTTGCGAATA CTGTAAAGTT	240
ATTCGTCGTA ATGGTCGTGT TATGGTAATT TGCCAGCAA ATCCAAAACA CAAACAACGT	300
CAAGGATAAG ATAGAAAGGA GAAACATGG CTCGTATTGC TGGAGTTGAT ATTCCAAATG	360
ACAAACGCGT AGTAATCTCA TTGACTTATG TTTATGGTAT CGGACTTGCA ACATCTAAGA	420
AAATTTTGGC TGCTGCTGGA ATCTCAGAAG ATGTTCTGTG ACGTGATCTT ACATCAGATC	480

877

AAGAAGATGC TATCCGTCGT GAAGTGGATG CAATCAAAGT TGAAGGTGAC CTCGTCGTG	540
AAGTAAACTT GAACATCAAA CGTTTGATGG AAATCGGTTC ATACCGTGGT ATCCGTCACC	600
GTCGTGGACT TCCTGTCCGT GGACAAAACA CTA AAAACAA CGCCCGCACT CGTAAAGGTA	660
AAGCTGTTGC GATTGCTGGT AAGAAAAAT AATATAGGAG GTAAAAGTCT TGGCTAAACC	720
AACACGTAAA CGTCGTGTGA AAAAGAATAT CGAATCTGGT ATTGCTCATA TTCACGCTAC	780
ATTTAATAAC ACTATTGTTA TGATTACTGA TGTGCATGGT AATGCAATTG CTTGGTCATC	840
AGCTGGTGCT CTTGGTTTCA AAGGTTCTCG TAAATCTACA CCATTCGCTG CTCAAATGGC	900
TTCTGAAGCT GCTGCTAAAT CTGCACAAGA ACACGGTCTT AAATCAGTTG AAGTTACTGT	960
AAAAGGTCCA GGTTCGTGGT GTGAGTCAGC TATTCGTGCG CTTGTGCGG CTGGTCTTGA	1020
AGTAACAGCA ATTCGTGATG TGACTCCAGT GCCACACAAT GGTGCTCGTC CTCCAAAACG	1080
TCGCCGTGTA TAATCATCGC ATTACACTGC TTTTCGTTA AGAGGGAGTA ACTAAATGAT	1140
CGAGTTTGAA AAACCAAATA TAACAAAAT TGATGAAAT AAAGATTATG GCAAGTTTGT	1200
AATCGAACCA CTTGAACGTG GCTACGGTAC AACTCTTGGT AACTCTCTTC GTCGTGTACT	1260
TCTAGCTTCT CTACCAGGAG CAGCTGTGAC ATCTATCAAC ATTGATGGTG TGTTACATGA	1320
GTTTGACACA GTTCCAGGTG TTCGTGAAGA CGTGATGCAA ATCATTCTGA ACATTAAAGG	1380
AATTGCAGTG AAATCGTACG TTGAAGACGA AAAATCATC GAACTGGATG TTGAAGGTCC	1440
TGCTGAAGTA ACAGCTGGTG ACATTTTGAC AGATAGCGAT ATTGAAATG TAAATCCAGA	1500
TCATTATCTC TTTACAATCG GTGAAGGTTT TTCTCTAAAA GCGACTATGA CTGTTAACAG	1560
TGGTCGTGGA TATGTACCTG CTGATGAAAA TAAAAAGGAT AATGCACCAG TTGGAACACT	1620
TGCTGTAGAT TCTATTTATA CACCAGTTAC AAAAGTCAAC TATCAAGTGG AACCTGCTCG	1680
TGTAGGTAGC AATGATGGTT TCGACAAAT AACCCCTGAA ATCTTGACAA ATGGAACAAT	1740
TATTCAGAA GATGCTTTAG GGCTTTCAGC ACGTATTTG ACAGAACATC TTCATTTGTT	1800
TACAAATCTT ACTGAGATTG CTAAGTCAAC TGAAGTGATG AAAGAAGCTG ATACTGAATC	1860
TGACGACCGT ATTTTAGATC GTACGATTGA GGAAGTGGAC TTGTCTGTGC GTTCATACAA	1920
CTGTTTAAAA CGTGCCCGTA TCAATACTGT GCATGATTG ACAGAAAAAT CTGAAGCAGA	1980
GATGATGAAA GTACGAAATC TTGGACGCAA GACTTTGGAA GAAGTGAAAC TCAAACATC	2040
TGATTTGGGT CTTGGATTAA AAGATAAATA AAGGAGGAAT ACATGGCTTA CCGTAAACTA	2100
GGACGCACTA GCTCACAACG TAAAGCAATG CTTCCGATT TGACAACTGA CCTTTTGATC	2160
AACGAATCAA TCGTGACAAC TGAAGCTCGT GCTAAAGAAA TCCGTAAAAC TGTGAAAAA	2220

878

ATGATTACTC TAGGTAAACG TGGTGATTTG CATGCACGTC GTCAAGCAGC TGCTTTCGTA	2280
CGTAATGAAA TCGCATCTGA AAACATATGAT GAAGCAACTG ATAAGTACAC TTCTACTACA	2340
GCACTTCAAA AATTGTTCTC AGAAATCGCA CCTCGTTATG CTGAACGTAA CCGTGGATAC	2400
ACTCGTATCC TTAACACTGA ATCACGTCGT GGTGATGCAG CGCCAATGGC GATCATCGAA	2460
TTAGTATAAA ATCATCAATT TTGTTGAGTG TTATGATGAT GGAGTCTTGT GCTCTTAGTC	2520
TAGCTCTGGT CTACCGCTAG GATTTTCGGTC CTAGCGGGAA CACTCATCAT AAGTTGGGAT	2580
AGTAGACGCT TGTTTACGAA ATTGTTTTTT TCTTAAGAAC AACTTCGTAA GCAGGCGTTT	2640
TTGAGTATTT TCGTTAGAAT TATGCTATAC TATTTGAAAA GAATCCTGTT TAATGTTAAG	2700
GTTTCTTATT TTAAGAAGAA TTGGAGTTTA CTTATGAAAG CCATTATAAC TGTGTTGGT	2760
AAAGATAAAT CTGGAATTGT TGCAGGTGTT TCTGGTAAAA TTGCAGAATT AGGATTGAAT	2820
ATTGACGATA TCTCTCAAAC TGTCTTGGAT GAATATTTTA CGATGATGGC TGTGTATCT	2880
AGTGATGAAA AGCAAGATTT TACCTATCTT CGTAATGAAT TTGAAGCTTT TGGGCAAACT	2940
TTGAATGTAA AAATCAATAT TCAGAGTGCA GCGATTTTCG AAGCTATGTA TAATATCTAG	3000
GAGGTCATCA TGGATATTAG ACAAGTTACT GAAACCATCG CCATGATTGA GGAGCAAAAC	3060
TTCGATATTA GAACCATTAC CATGGGGATT TCTCTTTTGG ACTGTATCGA TCCAGATATC	3120
AATCGTGCTG CGGAGAAAAT CTATCAAAAA ATTACGACAA AGGCGGCTAA TTTAGTAGCT	3180
GTTGGTGATG AAATGCGGC TGAGTTGGGA ATTCCTATCG TTAATAAGCG TGTATCGGTG	3240
ACACCTATTT CTCTGATTGG GGCAGCGACA GATGCGACGG ACTACGTGGT TCTGGCAAAA	3300
GCGCTTGATA AGGCTGCGAA AGAGATTGGT GTGGACTTTA TTGGTGGTTT TTCTGCCTTA	3360
GTACAAAAAG GTTATCAAAA GGGAGATGAG ATTCTCATCA ATTCCATTCC TCGCGCTTTG	3420
GCTGAGACGG ATAAGGCTCTG CTCGTCAGTC AATATCGGCT CAACCAAGTC TGGTATTAAT	3480
ATGACGGCTG TGGCAGATAT GGGACGAATT ATCAAGGAAA CAGCAAATCT TTCAGATATG	3540
GGAGTGGCCA AGTTGGTTGT ATTCGCTAAT GCTGTTGAGG ACAATCCATT TATGGCGGGT	3600
GCCTTTCATG GTGTTGGGGA AGCAGATGTT ATCATCAATG TCGGAGTTTC TGGTCTGGT	3660
GTTGTGAAAC GTGCTTTGGA AAAAGTTCGT GGACAGAGCT TTGATGTAGT AGCCGAAACA	3720
GTTAAGAAAA CTGCCTTTAA AATCACTCGT ATCGGTCAAT TGGTTGGTCA AATGGCCAGT	3780
GAGAGACTGG GTGTGGAGTT TGGTATTGTG GACTTGAGTT TGGCACCAAC CCCTGCGGTT	3840
GGAGACTCTG TGGCACGTGT CCTTGAGGAA ATGGGGCTAG AAACAGTTGG CACGCATGGA	3900
ACGACGGCTG CCTTGGCCCT CTTGAACGAC CAAGTTAAAA AGGTTGGAGT GATGGCCTGC	3960
AACCAAGTCG GTGGTTTATC TGGTGCCTTT ATCCCTGTTT CTGAGGATGA AGGAATGATT	4020

879

GCTGCAGTGC AAAATGGCTC TCTTAATTTA GAAAACTAG AAGCTATGAC GGCTATCTGT	4080
TCTGTTGGAT TGGATATGAT TGCCATCCCA GAAGATACGC CTGCTGAAAC TATTGCGGCT	4140
ATGATTGCGG ATGAAGCAGC AATCGGTGTT ATCAACATGA AAACAACAGC TGTCGTATC	4200
ATTCCCAAAG GAAAAGAAGG CGATATGATT GAGTTTGGTG GTCTATTAGG AACTGCACCC	4260
GTTATGAAGG TTAATGGGGC TTCGTCTGTC GACTTCATCT CTCGCGGTGG ACAAATCCCA	4320
GCACCAATTC ATAGTTTTAA AAATTAAGAA AATAGGAGAA ATTTTAAGTT CTATTTAAGA	4380
TTAGACGTGT ATACTATAAT CATTAAATAA AGACCTCCTA ATATTATTG AACAGATAA	4440
CACTGAATTA GTTTGAATTT GATTTTCATC TAATATCTTT ATTTAATGAA CTCCTAACT	4500
TTTTCATAAT AATCTCCTTC AAAAGTCGCC TGTATGGGTG GCTTTTATTT TATCATTCAT	4560
GATATAATAG AAGCAAACGG AGGACGGAAG ATGGTAAAAG TACGATTGTA TTTGGTACGT	4620
CATGGCAAGA CCATGTTTAA CACGATTGGT CGCGCGCAAG GTTGGAGCGA TACTCCCTTA	4680
ACTGCTGAAG GTGAACGAGG GATTCAAGAG TTAGGAATCG GTTTGCGAGA ATCTGATCTA	4740
CAGTTTGAGC GTGCTTATTC GAGTGATTCT GGTCTACCA TTCAGACCAT GCGAATTATC	4800
CTTGAAGAAC TTGGCTTGCA GGGGGAAATC CCTTATCGCA TGGACAAGCG TATCAGAGAA	4860
TGGTGTTCG GTAGTTTGA TGGAGCCTAT GATGGCGATC TTTTCATGGG CATTATTCCT	4920
CGTATCTTTA ATGTGGACCA CGTTCACCAA TTGTCTTATG CTGAACTGGC TGAGGGCTTG	4980
GTAGAGGTCG ATACAGCTGG TTGGGCTGAA GGCTGGGAAA AACTCAGTGG CCGAATCAAG	5040
GAAGGCTTTG AAATGATTGC AAAAGAAATG GAAGATCAAG GTGGAGGTAA CGCCCTTGTT	5100
GTCAGCCATG GAATGACTAT TGGAAACATT GTTATCTGA TTAATGGCAT GCATCCGCAT	5160
GGTCTGGATA ATGGTAGCGT GACAATCCTT GAATATGAGG ACGGCCAGTT TAGGGTTGAA	5220
GTTGTCCGTG ACCGTAGTTA CCGAGAGCTA GGACGTGAGA AGATGGAAGA AGGCTCTATT	5280
TAATCAGTCT AGACTTGCTT GCCATGAGCT AGGGATTGA TAAGAATATC AAGATAAGAA	5340
AAACAGCCG AGGGCACTCC TTTCGGCTGT TTTTGATGTG GAAAACTAAA GTGTAATGCT	5400
ATTGCTTTTA GAGATTTTCA TAAACAAGAG CAAGGAACCT ACTGTTAGAA CAGTCAGGAT	5460
AGTTGACAAG GTTGCGGCTA CACCGTAATT TCCTCTGAGA ACCTCTGTAT AAATAGCTAC	5520
AGTCATTGTT CTTGTTTTGA CATTGTAGAG GAGGATAGAA GTAGAGAGTT TTGAAATCAT	5580
TGTGACTCAA GATAAGATGG CTCCAGAAAT GATACCAGAT AGCATCATTG GAGTTGTAAT	5640
CTTAGCAAAG GTATTGAGAC GACTACTTCC TAAGCTTTCA GCAGCTTCTT CAATACTTGG	5700
TGCTATTTGT TGTAAGCTAG CAACAGATGA CGGAATAGTA TAAGGTAATC TTCTGGCAGA	5760

880

TAGAGACATA ATCAAGATGA AAGCAGTCCC TGTAATCATA AGAAATCCAC TTCCAAATAG	5820
ACCAGTATTG AAGGAAGAAA TGAAGGCAAT CCCTAGAACG GTTCCTGGTA CAATATAAGG	5880
TACCATACTG AGGCTGTCAA TTAAGTTTGT AAACAAATTC CGTTTTCTAA CGGCTAGGTA	5940
GGAGATAAAT GTCGCAAATA GAACAAC TAG AACTAAGGCA ATCAAAGGGA TACGAATGGT	6000
ATTGAAAATA GCAGATCCCA TACGATGGAA AGCTACCTTG TAACTGTTTG GAGAATAACC	6060
TTTAACAGAT ACCATACCTG ATGTTTTTAG GAAAGAGGTA TAAATTAAGT AGATTTGAGG	6120
TAAACAGAG ATAAAGATAA TTCCGTAGAC TGTGTCATAA ATGGCAGCCA TTTTTCCTTT	6180
TGTAGTTTTT TTAGGCTCAA TTGGATGGAG CAGATTCTAG CTGAAACTGT AGCGGTTTGC	6240
AATGTGTTTT TGGATAAGGA AAATTGCCAA GGCAATGATA ATCGCCATAA TTGCAAAAGC	6300
AGAATTTCTT CCAACCTCGC TAATAAATTG GGTATAAATC AGGACAGGGA AAGTCCGATA	6360
CCCTTCGCCA ATCAACATAG GCGTTCCAAA GTCTGAGAAT GCTCTCATAA ATACAAGCAA	6420
GGAGCTGCTA GTAAGGTTGG AACTAGGAGA GGTAAACAA CCGTTACGAT AGGTTTAAAT	6480
CCGAAGGACC CCATGCTTTC AGCTGCTTCA AGTAGAGAAT TGTCATACT GTTCATTGTT	6540
CCAGCAACAT ATAGAAATAC CAGTGGGAAT AGTTGCAGTG TAAAGACAAG TACAATTCCT	6600
TTGAATCAAT AAATATCGAT AGCTGGAAGA TAAAGGGCAT TTGTCAAAA TTTAGTGATC	6660
ACCTCATTTT GTCCTAGCAA GAGAACCAG GAGTAGGCTC CTACGAAAGG AGCTGACATG	6720
GAAGCAATGA TAATCAATAT TTGTAGAAAT TTCTTCCCCT TGAAGTCATA CATAGAGAAG	6780
AGATAAGCTA ATAGGGTTCC TACAAC TAAG GAAGTGATAG TAGCGGTAAT GGAAACCTTC	6840
AAACTGTTGA CTAGTGTCTC AGAGTAGTAG GCTTTACTAA AGAAAGTGAC AAAATTAGCT	6900
AGTGAGAATT GTCCTTCATG TATAAGTGCT TGCTTGAGCA CGGTAACGAT AGGATAAACC	6960
AGAAAGATAG GATAGGTAAG AAAGAGGAAG AAAGAGGAAA CTGTCCAAAT ATTTAGTTTT	7020
TTACGTTCCA TGGTTGACTC CTTTTATCAG GTTTTGGGAA CCATCTGCAG AAAAGATGTT	7080
TAATTTTTGC GTATTGATTC GTAGACGAAT ACGATTGCCT TTTTGTAGAT CTTCTTCAAA	7140
AGTTGATTCT TCACTAATT GAATTTTTGA GGCAAAACCT GTCTCAATGA AATAATCCGT	7200
ATTTAGTCCA AGATAGACGC TATCTCTAAT AGTTCTTCA ATATCTCCAG ATTCATCTTT	7260
GATAAACTCT TCGGGACGAA TGCTTACATG AATAGCTTGC TCCTCAACCT GATCAAGAGC	7320
TGGCATTCTGA AGGGCATAGC CATCTGAAA GACCATATAA GCGCCGTCGC TCCGTTTTTC	7380
AAGATTGGCA GGGATAATAT TTGTGCGTCC GATAAAGGTT GCCACAACT CATTAGCTGG	7440
TTTATGATAG AGTTCTTTTG GTCGGCCGAT TTGTTGGATC ACCCATCTT TCATAACAGC	7500
AATTTGGTCT GAAATAGCCA TGGCTTCTTC TTGGTCGTGG GTTACATAAA CAGTTGTAAT	7560



881

TCCCACCTTCG TGTGGATTT CTCGGATGGC TTGACGCATA TCCAAGCGAA GTTTGGCCTC	7620
CAGATTACTA AGTGGCTCGT CCATGAGGAG AACACTTGGA TTAACCGCTA AGGCGCATGC	7680
CAAGGTGACA CGTTGTTGTT GTCCACCACT GAGTTTATCG GGCTTTCGAT CCGCATATTG	7740
AGCAATTTGC ATGAGTTCAA GATACTTGTT GGTCTGTTGA ATCAATTCCT CTTTGGAAAC	7800
CTTCTTTTGC ATAAGACCAA AAGCAACGTT GTCTCGGACA GTCAAATGTG GGAAAATAGC	7860
GTAGTTTGG AAAACCATCC CGATATTGCG TTTGCTGGGT TCCATATTAT TGATTTTGT	7920
ATCATCGAAG TAAATTCCTC CACCTTCGAT ACTGTTGAAA CCTGCAATCA TACGAAGAAG	7980
GGTCGTTTTT CCACATCCTG AAGCTCCAAG AAGGGTAAAG AGACTTCCTT TTGGAATTGT	8040
AATGTTCAAA TTCTCAATAA CAGGGACATC GTGGTAGATT TTTTGGCGT TAATAATTTT	8100
GATCTCACTC ATAGTGAACC TCTTTTACTG TTTAGATTGG ATATCTGTAA AGACTTCGTT	8160
GTATTTCTTA ACGATATCTG ATTTATTCTT GATGACATAA TCATAATCTT CAGTGAGTGT	8220
TTTGATTTTG TCAATTGGTT TCATGTTTTT GCTTGTTTTA GCATTTTAC GAACAGGACG	8280
GTTAGTAGTG GTTGTAACAA GTGTATCTTG TACTTCTTGA GAGATAATAA AATCGATAAA	8340
TTTCTTGGCA TTTTCCATAT TTTTAGATTT TTTAACCATA GCAGCACTAG CAGGTAGGAA	8400
GACGGTTCCT TCTTTTGGAT AGACTACCTT AATGTTAGCT CCGTCATTTA AGAGTTTAAC	8460
TGCTGGATCT TCATAAGAGA GACCAACAGC CATTTCTCCA TCAGCGACTA CTTTATAGAC	8520
ACTAGATGAA CTTGAACCGA TTTTACCATC AATAAGTGTG AAAAGATCTT TTACATAAGA	8580
CCAAGCCTTA TCATCTTTGT AACCACCTTG AGCTTGTAGC ATATTTGTTA ATTGAGCAAA	8640
GGCGCTAGAA GAGTTTGCTG GGTGAGCAGT TGCGATTTTT CCTTTTAGTT CAGGTTTGAA	8700
AAGATCGTTA TATCCTTCGA TGTTCATGCC TTTAGTTAAA TCAGGGTTGA CGATTAAAC	8760
ACTACCATCT AGTGATAAG GAGTAGAGTA GCCAGTTGTG TTTTGATATT CTTTGATAAC	8820
ATTATCATT TCTTTTGAAG TATAGTTTTC AAAGACTTCT CCGTGGGTAG TATATTGTGT	8880
ATAAGAACCA CCAAAGATAA CATCAGCTAC AGGAACTTCT TTTTCTGACT CTAGTTTTTT	8940
GAAAAGTTCT CCAGTACCAG CTTGAATCAG TTCTACTTTG ATACCATATT TTTCTTCAAA	9000
GGCAGGAATA GTTGCTCCAA TTAAGCCCTC TGAGTTTGGT GAATAAACGA CTAGCGAACC	9060
GCCGTCTCCT TTATCAGATG AACTGTCATC GGCAGATTCA TTAGAAGAAC AAGCAGCATA	9120
ATACATCCAT TTCTTTTCA TGATGGATAC CTCCGTTGTG TTATTTAAGT TTATTTTAAA	9180
ACAATGTAAG CGTTTTTAAA ACATACAATT CTATTCTATA GTGTATTGAA TCTATAACAG	9240
TACACTTTGA CTGCTAAAAT ATTTCTATAA ATTAATTTGA CTTTCCTGAT AGAGATGTTT	9300

882

ACATCTTATT TCAATTCAC T ATATTAGAGT AAAATTCTCT AAAAAAGAA GAATAGCCTA	9360
TTTACTATT CTTCTGAGTG ATTTCAATTC CTTTGGGGAA ATATGGAGAT ACTTTTAA	9420
TCCTGACAAA TGGTTGTTTC TTTTCTAAA TCGGTGATAC TGTATCGGAG AATGCGCGTG	9480
AGGTCACAAA GGCTGCGATA GAGCTTCTAT GGAGAATTTC TTTTGGAGA GATTTTAA	9540
AGGAATGAGA CATCCGCTAC CTCCTTGAA GGTTTTG	9578

## (2) INFORMATION FOR SEQ ID NO: 128:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

CGGGCTGTTG TGACGATTCT TATTTCTATC TGTGTTATCT TTTGGGAAC TATTTTGGGT	60
GTTGTCTTGG CTTTGGGCA ACGTTCAAAG TTAAACCGC TTGTTTGGTT GGCCAAC TTG	120
TACGTTTGGG TTTTCCGTGG GACACCGATG ATGGTTCAAA TTATGATTGC CTTTGCTCTT	180
ATGCATATCA ATGCTCCGAC TATTCAGATT GGAATTTTAG GTGTTGATTT TTCGCGTCTG	240
ATTCCAGGGA TTTTGATTAT CTCTATGAAT ACTGGTGCTT ATGTTTCGGA GACTGTTCTG	300
GCCGGAATCA ATGCGGTTCC AAAAGGTCAG CTAGAAGCGG CTTATTCGCT AGGGATTCTG	360
CCTAAAAATG CGATGCGTTA TGTGATTTTG CCACAAGCAG TCAAAAATAT CTGCGCAGCA	420
TTGGGGAACG AATTATCAC CATTATCAAG GACAGCTCCC TCTTATCAGC TATTGGGGTC	480
ATGGAGTTGT GGAATGGGGC TACAACAGTT TCTACAACAA CCTATCTACC TTAAACACCA	540
CTTTTATTTG CAGCATTTTA CTACTTGATT ATGACCTCTA TTCTGACAGT AGCCTTGAAA	600
GCTTTTGAAA AACATATGGG ACAAGGAGAT AAGAAATAAT GACAGAAACC TTGATAAAAA	660
TTGAAAATTT ACATAAATCC TTTGGAAAGA ATGAAGTATT GAAGGGCATC AACCTCGAGA	720
TTAAAAGAGG AGAAGTTGTC GTTATCATCG GTCCTTCAGG GAGCGGGAAA TCTACCTTGC	780
TTGCTCTAT GAATTTGTTG GAAGAAGCAA CCAAGGGGAA GGTTATCTTT GAGGGAGTCG	840
ATATTACGGA CAAGAAGAAT GACCTGTTTG CCATGCGTGA GAAGATGGGC ATGGTTTTC	900
AACAATTCAA TCTCTTCTCT AATATGACTG TGATGGAAAA TATCACCTTG TCCCCTATCA	960
AGACCAAAGG TGACAGTAAG GCCGTGTCAG AGAAAAGAGC TCAGGAAC TTGGAAAAAG	1020
TTGGTTTGCC AGATAAGGCA GACGCTTATC CACAGAGTTT GTCAGGTGGC CAGCAACAGC	1080
GGATTGCCAT CGCGCGTGGG TTGGCTATGG AACCAGATGT TTTGCTCTTT GACGAGCCAA	1140

883

CTTCAGCCCT AGATCCTGAG ATGGTTGGAG AAGTTCTGGC TGTATGCAA GATCTAGCCA	1200
AGTCAGGAAT GACCATGGTT ATCGTAACAC ATGAGATGGG ATTTGCCCCG GAGGTGGCAG	1260
ATCGTGTCAT CTTTATGGCA GACGGTGTGG TTGTTGAAGA CGGAACACCT GAGCAGATTT	1320
TTGAACAAAC CCAAGGACAA AGGACTAAAG ACTTCTTGAG TAAGGTTTTA TAAGTTAGCT	1380
TTGTTTAGCT ATTTGTAGCC AGCTTTAAAC GTTAAAGAGA AGATTAGTGA AAAGCTCAAC	1440
CAGAGCTTTT TCTTATAGTT TAAAGCTATA GGATTGCCTA GGAAAGAAGT GTTAGAGCTA	1500
CATTGTATTT TTTGGTATAA TTAAGATAT TTGTAAGAAA AGAGAAGTGA TATGACACAG	1560
ATTATTGATG GGAAAGCTTT AGCGGCCAAA TTGCAGGGGC AGTTGGCTGA AAAGACTGCA	1620
AAATTAAAGG AAGAAACAGG TCTAGTGCCT GGTTCGGTAG TGATTTTGGT TGGGGACAAT	1680
CCAGCCAGCC AAGTCTACGT TCGCAACAAG GAGAGGTCAG CCCTTGCGGC TGGTTTCCGT	1740
AGCGAAGTAG TACGGGTTCC AGAGACCATT ACTCAAGAGG AATTGTTAGA CCTGATTGCT	1800
AAATACAATC AGGATCCAGC TTGGCATGGG ATTTTGGTTC AGTTGCCATT ACCAAAACAC	1860
ATTGATGAAG AGGCGGTTCT ATTGGCTATT GACCCAGAAA AGGATGTGGA TGGTTTCCAT	1920
CCTCTAAACA TGGGGCGTCT TTGGTCTGGT CATCCAGTCA TGATTCCTTC GACACCGGCA	1980
GGAATTATCG AAATGTTCCA TGAATATGGG ATTGACTTGG AAGGTAAAAA TGCAGTCGTC	2040
ATCGGTCGAT CCAATATTGT CGGAAAACCT ATGGCCCAGC TTCTTTTGGC AAAGAATGCA	2100
ACAGTAACCT TGACTCACTC ACGTACTCAT AATCTTTCCA AGGTGGCTGC AAAAGCAGAT	2160
ATTCTGGTTG TTGCAATCGG TCGTGCCAAG TTTGTGACTG CTGACTTTGT CAAACCAGGT	2220
GCGGTAGTCA TTGACGTTGG GATGAACCGC GATGAAAATG GTAAGCTCTG TGGGCATGTT	2280
GATTATGAGG CGGTTGCCCC ACTTGCTAGC CACATTACGC CAGTCCCTGG AGGTGTCGGT	2340
CCTATGACCA TTACTATGCT GATGGAGCAA ACCTATCAGG CAGCACTTAG GACATTGGAT	2400
AGAAAATAAG ATAAAAATTT TCTGAGGAAA GTGTATTTTC TATAGCTATA TCTAAATGA	2460
TAGAAATGAA TATTAAATTT TAGAAATAAG TTTATAAAG GAGGTTTGCG CCTCCTTTTT	2520
GTTGTATAAT GGAGTGAGGT GATTAGATGA TTTTAAAAAT TTATAATGGG GAATATAGTT	2580
TACAATGGGA TGGAATATAC TACTTAGCAC TAATTGATTA TCCAAATATT CAAGAGTGGG	2640
AATTAGAAAA AATTGCTAAA TTTATAGCTT ACGAAAAACT TCATAAACGT CAAACAAGTA	2700
TTGAGTGTGC TGATTCTTGT TTAATAAAG AAATTTTAGA TTACATCTGT CAGCATCCCT	2760
TTCTGCCACC ATTTACTCCT ACAGATAAAA GAGTAGCCTC GACTTATGAC CTACATAAGA	2820
GGTTAGTGAC TTCAGACTAC TGTAGTCATA CTACGACTAT AGATGCAGCG ATTTCTATTT	2880

884

TTAAACTGG TCGTCTTTTA TCTGCTGTGA AAGCCTTTGG GCGAGATGCT GAGGAGTTGG	2940
TTTTGGATAG TCGAAATGCT GCATCTGATC CGATAGATTA TTTTGACTAT GTCATGTTAG	3000
GGTGGTCAAA TACAAGTTCT GGTATCGAT TGGCGATGGA GCGTTTATTA GGTGAGCTC	3060
CTTCAGAGAA AGAATTACAA GACAAGTTTA TTCCTGGAGT AAGTTTTCAT TTTATCTATA	3120
CAGATTTGAT TAAAGTTCCT GGTATATTTT TTGATGGTTA CATTGCTGTA AAAATTAAGG	3180
ACATGCTTAA TTTATTAAGT GAGTTGTATA TTTGCATTAT TCCAACTCAT AATAAGAGCC	3240
AATTTGAAAA TATTATCCA ACCAAAATAC AAGATAGGGT GTATTATCTT GACTATGCTG	3300
GAGAAGACTT AGAAGAGTGG ACTAAGAAAG TCTATCAAGT TGTTTTAAAA CAATCAGATA	3360
AAGGATAGTT GAGGAAAAAA CGATGAAAGT GATTGATCAA ACCTTACTAG AAAAAGTCAT	3420
TATTGAACGT TCTGTACAA GTCATAAAGG AGACTACGGT CGTCTGCTGT TGCTTGGTGG	3480
GACTTATCCT TATGGTGGTG CCATCATCAT GGCTGCTTTA GCAGCTGTAA AAAGCGGTGC	3540
AGGATTGGTA ACCGTTGGAA CGGACAGGGA AAATATCCCT GCTCTACACA GCCATTTGCC	3600
TGAGGCTATG GCCTTTTCTC TGCAAGATCA GTAATTGTTA CAAGAGCAAT TGGAGAAGGC	3660
AGAAGTTGTC TTGCTGGGGC CTGGTTTACG AGACCATACG TTTGGAGAAA ATCTTGTAAG	3720
ACAGGTCTTT GCTAGCTTAA AAAAGAATCA GATTTTGATT GTAGATGGAG GGGCCTTAAC	3780
CATCCTTGCT AGGACAAGTT TGTTGTTTCC ATCTAACCAG CTTATCTTAA CTCCCCACCA	3840
AAAAGAATGG GAAAACTGT CTGGTATTGC TATTGAAAAG CAAAACGAAG GTACAACATC	3900
TAGTGCCCTG ACTTCTTTCC CTCAAGGAAC AATTTTGGTA GAGAAAGGTC CAGCTACTCG	3960
TATTTGGCAA GTTGGCCAGT CTGATTATTA CCAGTTAAAG GTTGGCGGTC CCTATCAGGC	4020
GACTGGTGGT ATGGGTGATA CACTGGCTGG AATGATTGCA GGATTTGCAG GCCAATTTCC	4080
ACAGGCCAGT CTCTACGAAC GTGTGGCAGT AGCAACCCAT CTTCAATTCAG CCATAGCCCA	4140
AGAACTATCT CAAGAAAATT ATGTGGTCTT GCCGACGGAA ATTAGTAATT GTCTTCCTAA	4200
AGTAATGAAA AGATATGTCT AAAATAGTTA GACAAAAAAT GTTGATAATT TGTATCATT	4260
TTCTTAATTC AAAAAAAGC AACGTTTAGT ATTCTTCTTG CTAAGAACT AAATTTGTTC	4320
GTTTTTTTAC TCTTGTAAT CTATTTTGT TAGAGTTGAT TTGGTTTACA TCCGTACTTA	4380
AATTGATTG TTAGAGCTCT ACTTTTATTA AAAAAATCA ATTTCAAGGA TAAATAAGCA	4440
GTATTCTAAA GGTACTTTTA GATGAAATAA AAGCCTTTAC ATGGTATAAT AGAGGTAGCT	4500
CTTTAATGGA GGTGTTTGAG TGGAAAATCT GAAGAAAATG GCAGGTATCA CGGCTGTGA	4560
ATTTATCAAG GATGGGATGG TTGTAGGGCT AGGAACAGGT TCTACTGCCT ATTATTTGT	4620
CGAAGAAATC GGTCGTCGAA TCAAGGAAGA AGCCTTGAG ATTACAGCTG TGACGACTTC	4680

885

TAGTGTGACC AGTAAACAGG CTGAAGGGCT CAATATCCCG CTCAAGTCTA TTGACCAAGT	4740
AGACTTTGTC GATGTGACAG TCGACGGGGC GGATGAAGTG GATAGTCACT TTAATGGAAT	4800
CAAAGGCGGT GGTGGTGCCC TTCTCATGGA AAAGGTGGTC GCAACACCAT CAAAAGAATA	4860
CATTTGGGTG GTGGATGAAA GCAAGCTGGT CGAAAACTA GGTGCTTTTA AATTGCCAGT	4920
AGAAGTGGTT CAGTATGGTG CAGAGCAGGT CTTTCGTCAT TTTGAACGAG CTGGCTACAA	4980
ACCAAGTTTC CGTGAAAAAG ACGGCCAACG TTTTGTGACC GATATGCAGA ATTTTATCAT	5040
TGACCTCGCC TTGGATGTCA TTGAAAATCC AATTGCTTTT GGACAAGAAT TGGACCATGT	5100
CGTTGGTGTT GTGGAGCATG GTTTATTCAA CCAATGGTG GATAAGGTAA TCGTTGCTGG	5160
ACCAGATGGA GTTCAGATTT CAACTTCAAA AAAAGGAAAA TAGAAGGGGG CATAAGATGT	5220
CTAAATTTAA TCGTATTCAT TTGGTGGTAC TGGATTCTGT AGGAATCGGT GCAGCACCAG	5280
ATGCTAATAA CTTTGTCAAT GCAGGGGTTT CAGATGGAGC TTCTGACACA CTGGGACACA	5340
TTTCAAAAAC AGTTGGTTTG AATGTCCCAA ACATGGCTAA AATAGGTCTT GGAAATATTC	5400
CTCGTGAAC TCCTCTTAAG ACTGTAGCAG CTGAAAGCAA TCCAACCTGA TATGCAACAA	5460
AATTAGAGGA AGTATCTCTT GGTAAGGATA CTATGACTGG ACACTGGGAA ATCATGGGAC	5520
TCAACATTAC TGAGCCTTTC GATACTTTCT GGAACGGATT CCCAGAAGAA ATCCTGACAA	5580
AAATCGAAGA ATTCTCAGGA CGCAAGGTTA TTCGTGAAGC CAACAAACCT TATTCAGGAA	5640
CGGCTGTTAT CTATGATTTT GGACCACGTC AGATGGAAAC TGGAGAGTTG ATTATCTATA	5700
CTTCAGCTGA CCCTGTTTGT CAGATTGCTG CCCACGAAGA CATTATTCCT TTGGATGAAT	5760
TGTACCGTAT CTGTGAATAC GCTCGTTCGA TTACCCTTGA GCGTCCTGCC CTTCTTGGTC	5820
GCATCATTGC TCGCCCTTAT GTAGGTGAAC CAGGTAACTT CACTCGTACG GCAAACCGTC	5880
GTGACTTGGC TGTATCTCCA TTTTCCCAA CTGTTTGGGA TAAATTGAAT GAGGCTGGTA	5940
TCGATACTTA TGCTGTGGGT AAAATCAACG ATATCTTTAA CGGTGCTGGT ATCAACCATG	6000
ACATGGGTCA CAACAAGTCA AATAGTCATG GAATTGATAC ACTATTGAAG ACTATGGGAC	6060
TTGCTGAGTT TGAAAAAGGA TTCTCATTC AAAACCTAGT TGAATTTGAT GCCCTTTACG	6120
GCCATCGTCG TAATGCTCAC GGTACCCTG ATTGCTTGCA TGAGTTTGAT GAACGCTTAC	6180
CTGAAATTAT CGCAGCTATG AGAGAGAATG ACCTTCTCTT GATTACTGCG GACCATGGAA	6240
ATGACCCAAC GTATGCAGGA ACGGATCACA CTCGGGAATA TATTCATTG TTGGCCTATA	6300
GCCCTGCCTT TAAAGGAAAT GGTCTCATTC CAGTAGGACA TTTTGCAGAT ATTTCAGCGA	6360
CTGTTGCCGA TAACTTTGGT GTGGAACTG CTATGATTGG GGAAAGTTTC TTAGATAAAT	6420

886

TGGTATAAGA TGACGCGCTA TGCTTTGCTG GTGAGAGGTA TCAATGTTGG TGGTAAGAAT	6480
AAGGTCGTCA TGGCGGAGCT TCGTCAAGAA TTGACAAACT TGGGACTGGA AAAGGTTGAG	6540
AGCTACATCA ATAGTGGCAA TATTTTCTTT ACTTCGATAG ATTCCAAAGC CCAATTGGTT	6600
GAAAAGCTAG AGACTTTCTT TGCAGTCCAT TATCCATTTA TTCAGAGCTT TTCTTTACTG	6660
AGTCTAGAGG ACTTTGAGGC GGAAGTTGAA AATCTACCAG CTTGGTGGAG CAGAGACTTG	6720
GCACGAAAAG ATTTTCTCTT TTACACTGAG GTTTGGATG TGGACCAAGT CATCGCGACA	6780
GTTGAAAGTT TAGAGCTGAA AGATGAAGTG CTTTATTTTG GAAAAGTTGG GATTTTCTGG	6840
GGGAAATTTT CTGAAGAATC CTATTCTAAG ACTGCCTATC ATAAGTACTT GCTGAAGGTG	6900
CCTTTCTACC GCCACATTAC TATTCGTAAT GCTAAAACCT TTGACAAAAT TGGTCAAATG	6960
CTAAAAAAT AATAAAGGAG ACACACAATG ACATTTTTAA AAAAAATCCA TGAACTGCT	7020
ACTTTCCTGA AAGAAAAGG AATTGCAGCC CCTGAGTTCG GTCTAATCCT TGGATCAGGA	7080
CTTGGAAGT TGGCAGAAGA AATCGAAAAT CCAGTTGTAG TAGACTATGC TGAGATTCCA	7140
AAGTGGGGCC GTTCAACAGT AGTCGGTCAT GCTGGTAAAT TGGTATATGG TGAAGTGGCA	7200
GGTCGCAAGG TCTTGGCTCT TCAAGGGCGT TTCCATTTCT ATGAAGGGAA TCCTCTGGAA	7260
GTGGTGACTT TCCCAGTTCG TGTGATGAAA GTTCTTGGAT GTGAAGGTGT TATTGTAACC	7320
AATGCAGCTG GCGGTATCGG ATTTGGTCTT GGTACCTTGA TGGCTATCTC AGACCATATC	7380
AACATGACGG GGCAAAATCC ATTGATGGGT GAAAAGTTGG ATGACTTTGG CCCACGTTTC	7440
CCAGATATGT CTAGGGCCTA CACACCAGAA TACCGTGCCA CTGCCCATGA AGTGGCTAAA	7500
AAACTTAATA TCAAGCTTGA TGAAGGTGTC TATATCGGAG TTAAGTGTCC GACTTATGAA	7560
ACACCAGCAG AAATTCGTTT CTATAAGACA CTGGGAGCAG ATGCAGTTGG TATGTCTACG	7620
GTTCTGAAG TTATCGTGGC AGCCCACTCT GGCTTGAAAG TTCTGGGAAT TTCATGTATC	7680
ACTAACTTTG CGGCCGGTTT CCAAGAAGAA CTCAATCACG AAGAAGTTGT AGAAGTGAAT	7740
GAACGTGTTA AAGGTGATTT CAAAGGCTTG CTAAAGCGA TTCTTGCTGA ATTGTAAGAA	7800
AAAAGATTTA AAAGGGGGAG TGCCTCTGTT TTTTCAGGAT TGAAGTCTTA TCCGGATTAA	7860
AGAAGAAACA GAGGAATACT ATGAGCTTCT TCCTGCTCTT ATAAGTAAA GAAGCGGAAG	7920
AATAGGTATG TCTGATCTGA TAGCCAGCAT TGTGAAAGAC AAGATTCTAG GATACTAGCA	7980
TTAGCTTCTT AGCCAAGCAG ACTAGTATGA TAAGGAGAGA TGAGAATGAA TTGACTTTCT	8040
GAATTTCTCA GTCTTATCAT ATATAGCACA ATGAGATTTT GCTTGAGTCT GCTTGTAAT	8100
AAACGAAAAG AAAGATAAGA AATAATGAAA ATTGGTCAAC GAATTATGCG CTTTGGCATA	8160
AAAAATTAAG TATCGGAGTT GTATCTGTTG TAGTCGGCTT TGATTTCTAG CTCCAGCTGG	8220

887

AATTTTCAGCC AATGAAGTAA AGCAAGATGT AACATCTGAA GTGGTAATAG GTGTGCTAGA	8280
TTCTAAGGAG GAATTGAAAG AGTCAGAAAA TGATGCTCCA AAAGTAGAAA CTCCTCTTAG	8340
AGAGGAGCCA AGACTAGCTC CTCAAACGCT TCCGGAAGCA AGTGAAGTTC TTGAAAACAA	8400
AAGGGAAGAG TCAAAAGTAG AGATAACATA ACCAGCTCAA GCGGATGATA TCCGCAAGGT	8460
TGTTGGGGAA TTAGCCAAGG ATATAAGTAT TACTAAGTTG TATATGACAG GTCATTCTCT	8520
TGGATGTTAC CTAGCTCAGA TTGCAGCGGT TGAAGCTTAC CAAAAATATC CTGATTTTTA	8580
TAACCATGTA TTGAGGAAAG TGACAACCTT CAGTGCTCCT AAAGTGATTA CTTCCAGAAC	8640
TGTTTGAAT GCTAAGAATG GTTCTGGGA TGTGGTTTG GAAAGTCGTA AATTAGCTGT	8700
TAGTGGAAAA ATTAAGCATT ATGTGGTTGA TAATGACAAT GTTGTGACTC CCTTGATTCA	8760
TAATAATCGT GATATTGTTA CATTTACAGG TAATTCACGC TTAAACACC GTTCTCGTGG	8820
CTATTTTGAA AGTCCAATGA ATGATATTCC TAACTTAAT ATTGGTAAAC AAGCTACCTT	8880
GGATAAACAT GGTATCGTG ATCCGAAATT GGATAAAGTG CGATTCTTTA AGAAACAGGC	8940
TCTGCCTCGA TCTTCTAGTC AACCAAGCGC TGAACCAATG GAAAATATTG CCTCAGGAAA	9000
ACAGGTTACT CAAAGTTCGA CAGCTTTCGG AGGAGATGCT AGAAGAGCTG TGGATGGCAA	9060
AGTCGATGGT AACTATGGTC ACAATTCTGT CACTCATACA AACTTCCAAT CTAAGCCTTG	9120
GTGGCAAGTA GATTTGGCTA AAGAAGAAAC CATTCGCCAA ATCAATATTT ACAACCGAAC	9180
AGACACTGCC CAGGATAGAT TGGCAAACCT TGATGTCATT CTTTGTAGACA GTTCTGGTAA	9240
AGAAATTGAG TGAACCGTA TAACATCTCC TAAAGATGTG TCAGCACAAA TTACGATTAA	9300
CCATAAAAAA GCGCGCTATG TTCGGATTGA GCTAGAAGGC TATAATGCCC TCAGTCTTGC	9360
AGAAGTTGAA GTTTTCTGCT TTATAGCTAC GAATGCTGAA ACGGCGACAC AAGTTTCTAA	9420
GCCAGTTCAA CCAATCAGTC AGACTCCTGT GAAGGATAAA ACATTGACAA TTCAACACAG	9480
TGGAGCTTAC ATTGCCCCT ACTCCATAAC TTGGGAAGAA GTTCCAGTAG ATAAAGATGG	9540
AAACCAAGTT GTTCGTAGTC ATTCTTGGGA AGGAAGCGGT CGCAACCAGA CTGCAGGTTT	9600
TGTCCTCAAC CTCCCAATCA AAGAAAATAT GAGAAATCTG CGAGTTAAGA TTGAGAAAAA	9660
GACGGGCCTA CTATGGAATA GATGGCAAAC AATCTATGAA AACAGACCAA TTTTAGCTCA	9720
ACCCACCGT AAAATTACCC ATTGGGTAC GACATTGAAT TCCAAGGTGA GTGACGATGA	9780
TGTCTTGTA TCTGATGGTA GAATGACAGT TAGTTTGTCT AGTTTATAAG AAAGTACTAC	9840
CTGAGCTTGA ATAGGACTCA GGTAGCTCTC TATGAAAGAA CAAAATTAAT ACTCAATGAA	9900
AATCAAAGAG CAAACTAAGA AACTAGCCGC AGGTTGCTCA AAGCACTGCT TTGAGGTTGT	9960

888

AGATAAGACT GACGAAGTCA GTCACATATA TAATCCAAGG CGACGTTGAC GTGTTTGAA	10020
GAGATTTTCG AAGAGTATAA ACAGAAAGGT AGAGCGCGTG TTCTAATTG AACACGAGTA	10080
GAAAACTTTT CTA AAAACAA AAACGAAAGG ATGGGTAAAC TGTATTCGCT GAACTGAATA	10140
CGGGCGACTC TCCTCTAAAT CAAAATTAAG AAAGGAATTG ACCCCACCCT AAAAGTAGTG	10200
GGAAAAAGAT AGTTGATCTA GCGAGCATCG CTCACTGCGC CCAACTCCTA TTTTCCCTTC	10260
GCTTTTGTAT GGGTTTGTA TCTTCTCAA TATAAATAT AAAATAAAGA AAGGTAGAGC	10320
GTGTGTTTTG ATTTGAACAC GAGCGGAAAA CTCGGAAAAAT AGATAATCTG ACTGAAAAAT	10380
CAGGATTTCT CGTCAGGTTC CTAATTTTCA GTCGTTTTCT TCTCGCTCTT TGTATCATAA	10440
ATTATGTCTA TCCATATTGC TGCTCAGCAG GGTGAAATTG CTGATAAAAT TCTTCTTCCT	10500
GGGGATCCTC TTCGTGCTAA GTTTATTGCG GAGAATTTC TTGATGATGC TGTGTGTTTT	10560
AACGAAGTGC GTAACATGTT TGGTTACACT GGTACTTACA AGGGTCACTG TGTATCTGTC	10620
ATGGGAACTG GGATGGGAAT GCCATCTATT TCGATTATG CGCGTGAGTT AATCGTAGAC	10680
TACGGTGTGA AGAAATTGAT TCGTGTGGGA ACTGCAGGTT CTTTGAATGA AGAGGTTTAT	10740
GTTCGTGAAT TAGTTTTGGC GCAGGCGGCT GCAACCAACT CAAACATCGT TCGTAATGAC	10800
TGGCCACAGT ACGATTTTCC ACAAATTGCT AGCTTTGATT TGCTTGATAA AGCCTACCAT	10860
ATCGCCAAAA AACTTGGTAT GACTACTCAC GTTGGGAACG TTTTGTATC TGATGTCTTT	10920
TACTCAAATT ACTTTGAAAA GAATATCGAG CTTGCTAAAT GGGGAGTCAA GGCTGTGGAA	10980
ATGGAAGCAG CAGCTCTTTA CTATCTTGCT GCCCAATACC ATGTTGATGC GCTAGCTATC	11040
ATGACCATCT CTGATAGCTT GGTCAATCCA GACGAAGACA CAACTGCAGA AGAACGTCAA	11100
AATACCTTCA CTGATATGAT GAAGGTTGGT TTGGAAACCT TGATTGCAGA ATAATTATAG	11160
CCAAAAAGGG GCTCTTTGTC AACTGTAGTG GGTGAAAAA AAGCTAAGCT TGAGAAAGGA	11220
CAAATTTTCG CTTTCTTTT TTGATATTCA GGGCGATAAA AATCCGTTTT TTGAAGTTTT	11280
CAAAGTTCCG AAAACCAAAG GCATTGCGCT TGATAAGTTT GATGAGATTA TTGTCGCTT	11340
CCAGTTTGGC ATTAGAATAG TGATGTTGAA GGGCGTTGAC GATTTTCTCT TTGTTCTTTA	11400
GAAAGGTTTT AAAGACAGTC TGAAAAAGAG GATGAACCTG CTTGAGATTG TCCTCAATGA	11460
GTCCGAAAAA TTTCTCAGGG TCTTTGTTCT GAAAGTGAAA AAGTAAGAGT TGATAGATCT	11520
GATAGTGGTG TTTCAAGTCT TCTGAATAGC TTAAATCTT GTCAAGAATT TCTTTATTG	11580
TTAAGTGCAT GCGAAAAGTA GGGCGATAAA AACGTTTATC GCTsAATTTA CGACTATCCT	11640
GTTGGATGAG TTTCCAGTAA CGCTTGATAG CCTTGATTG ATGAGATTTT CGTTCAAAC	11700
GATTCATAAT TTGAACACGA AAACGACTCA TGGCAGGCT GAGATGTTGG ATAATATGGA	11760



889

AACGATCTAG AACGATTTTA GCACACGGAA AAAGCTGTTT AGCCAAGTCA TAGTAAGGAC 11820  
TAAACATATC CATCGTAATG ATTTTCACTT GACAACGAAC GGCTCTATCG TAGCGAAGAA 11880  
AGTGATTTTCG GATGACAGCT TGTGTTCTGC CTTCAAGAAC AGTGATAATA TTAAGATTAT 11940  
CAAAATCTTG CGCAATGAAA CTCATCTTTC CCTTAGTGAA GGCATACTCA TCCCAAGACA 12000  
TAATCTTTGG AAGCCGAGAA AAATCATGCT CAAAGTGAAA GTCATTGAGC TTGCGAATGA 12060  
CAGTTGAAGT TGAAATGGCC AGCTGATGGG CAATATCAGT CATAGAAATT TTTTCAATTA 12120  
ACTTTTGAGC AATTTTTCG TTGATGATAC GAGGGATTTG GTGATTTTTC TTTACCAGGG 12180  
GAGTCTCAGC AACCATCAT TTTGAASAGT GATAGCACTT GAAACGGCGT TTTCTAAGGA 12240  
GAATTCCTAGA AGGCATACCA GTTGTTCGA GGTAAGGGAT CTTAGACGGT TTTTGAAAGT 12300  
CATTTTCTT CATTAGACTT CCACAATCAG GGCAAGATGG AGCCTCATAA TCCAGCTTAG 12360  
CGATAATTTT TTTGTGGGTA TCCATATTGA TGATATCTAG AATCTTGATG TTTGGGTCTT 12420  
TAATATCGAG CAGTTTGTG ATAAAATGTA ATTGTTCCAT ATGATTCTTT CTAATGAGTT 12480  
GTTTTGTGCG TTTTCATTAT AGGTCATATG GGACTTTTTT TCTACACAAA AATAGGCTCC 12540  
ATAATATCTA TAGTGGATTT ACCCACTACA AATATTATAG AGCCCCAAAA GGAAGCCCTT 12600  
TATGAATTGT AGGACTTCCT TTTCTTATCC AGAAATTGAT CTAGCTCTCT CTGATTTTCA 12660  
AGAATAGTGA CTTTATGTGA ATATTCTTGG CAAAGTTTTT GGTAATTTTC TTTTGTAGTT 12720  
TTGCGGACGC CCATCCCAA GAATCCATCT GATAAACTCC CACTCAAAGC GTTCAGGGCA 12780  
ATCTACCGCC ATACTTTCTC TGACTTTTCC ACGGTATTTA AGATAACGCT TAAAGGCTCT 12840  
AAAGAGACAG GTCAATGGCG AAAAATTGAG AAAGATGATT TGGTCAGCTT CTTGCATTCC 12900  
TTCTTGGTAG TAGCACCAAG AATAATTACC ATCGATGACC CAAGCTTTAT GCTTGGTGAG 12960  
AAAGTTTTTT ATCTCGGTTA ACATCCATTC GCAGTCACTG TCTTGCCAAC CAGGTTGAAA 13020  
TTGGAGTGTG TCCATGTGCA GTTTTGAAT GGAGTAGTAG TTAGATAACT TTTCTGCTAT 13080  
AGTTGACTTA CCAGAACCAG AATATCCGAT AATTGCGATT TTCATTTTCT ACCTTTTCCT 13140  
ATTTGGAGAC AAAAAACAG CCTCTATGGA CTGTTTCTTA TTTAACAAGT TTAGCTGAAA 13200  
GACGAGCTTT ATCGCGGCTT GCTTTGTTTT TGTGAATCAA ACCTTTAGTT TCTGCTTTAT 13260  
CGATAGCTGA GCTAGCAGCA CGGAAAAGTT CTTGAGATGG GTTTGCTTCG AAAGCTTTTA 13320  
TAGCAGTACG CATAGCTGAT TTTTGAGCTG AGTTCTTTTC GATTCGTCTA ACGTTCAATT 13380  
CAGCGCGTTT GATAGCTGAT TTAATGTTTG CCAATGGTCT TACCTCCATA TTTACTAACT 13440

(2) INFORMATION FOR SEQ ID NO: 129:

890

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8512 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

CCTTTTTTCA AAAACTAGAT ACTAGTCTAT CAAAAGTAGG AAAGGGTTTC AAGAAAATTG	60
ATTGGAAAAT TTTTGAAAAT CATAGAACTA TTAGCTAATC CCTAGTATTG AAAAGACTGG	120
ATAGCTTCTT TCAGGTCATC TTGTAACTA TTTCTCTGGT CAAGTTGGAC ATAGACTTCC	180
ACCAGACAGG ATCTAAAGTT GGAATAATTG TAAAAATCCT CCTTTCTTC TATCGGAAAA	240
TCAACAGTTT TTATCCAAGA AGCTACTTGT TCTTGCTCCA ACTTCCCTTG TAAATAGGT	300
TCATAGATCA CTCTTGCTAA ACGCCAATCC TCATCATCTG TAAAGCGAAT CGACATTCTT	360
TTAAATAGTT GGCCAAGTAT ATCAAATACT TCATGAATC TGTTTTTAGG AAAGTCTGGA	420
TGACAAACCA CCTCTGTCAG TAAATCGGCT CCATGTGCAA AAGCGTGAAC CCAACCATAC	480
TGACTTGAGA AACCCTTGT ATCCTTTTCT TTTGAAAGAT AGTGCAAGCC TTGATTTAAA	540
AGGACATTAC GAATTTCTGG AGAAGGATTT CCCAAATGAT CAAACAACCA CTGGATTCTT	600
TCCTGGTTAT AATTGGTTT TTCTTCTGCT ATTTTCTTA GTAAATCTTG ATACATGGTC	660
AATACCTCTA CATTCTAGC AACTGTTCAA AAAGGCAGTC TTAAATGACT CAATATTGAA	720
TTCTCAATTA AATACAATCT GATATAAAAT GACGTAAATA ACTATCAATA CCAGTTCTAC	780
AGTAAGTTCA AATTTAACAT CACGACCTTC AACGACATTT TTGAAAATAG CTACAACATA	840
GACAAATAGA ATGACGCTTA ACAAGCCCAT AACATCATT CTAATAAATT TTTCTATTCC	900
CCTACTCTCC CAACTCAGCA CTATAGGAGA TAATCTGTC AACTGTGTCA GACAAGAATT	960
GGATGGTATC ACGGAGTGGT TTGTCTGTTG AAATATCAGC ACCGATAATC ATGGCTGACT	1020
CAAGTGGTGT CTTGCTACCA CCTGATTTGA GGAGATTGAG CCAGTCTTCA GCTCCAGTTT	1080
CAGAATGTTT TAGATGAAGG TAACCAGCAG TCGAGATAAC TAGTCTGCT GAGTAAGTGT	1140
AACTATACAA GCCCATATAG TAGTGAGCTT GGCGCATCCA AGTCAGAGTT GCATCATCGT	1200
CAATTTCAAT AGCATCTCCC CAGAAATCCG TCAAACTTC CTTCATAATG CTGTTGAGCT	1260
TGCTTGCTCC AAAGGTCTCC CCTTCTTCAA TCAATGTATA CACCTTACGC TGAAGGCGG	1320
CTTCCAAGAG GTGGGTGATA AAGTTATGGA AGTAGGTGTC TGTCAAGCGA TGAGCCAGAG	1380
CGAAGCGTTT TTGACGTGGG TCATTAGACT GGTTCCTCAA GTAATCACTG AGTAGCAATT	1440
CATTGAAGGT TGACGGTGCT TCAACATAGT AGGTCGACAT ATGGGCATTG AAGTAACTTT	1500

891

GATGATTGTC TGAAAAGATG AATTGACCAG AATGCCCGAT TTCATGAATC AAGGTATAGA	1560
CATCGCTCAA ACGGCCTGTC CAGCTCATGA GTACATAAGG GTGTACGCGA TATGGGTCCG	1620
CCGCATAACC ACCGGAATCC TTGCCACTGT TAGCAGCAAA GTCCACCCAG CGCTCTTCTT	1680
GGTAACGAGC AACTTCCTGA CAATATTCTT GCCCCAAAGG TTCTACCGAC TTCATGACCA	1740
AATCATAGGC ATCGTCAATA GTCACTTCAG GATTCAGGGC GCTGTCCAAG TCCAATTTC	1800
AGTCTGCAAA GGTCACTTT TCAAGACCAT TTACCTTGGC AACATGCTTG AGGTATCTCT	1860
GAGCGACTGG TGCAAAGTCC TTCATCATGA GGTCATCTG GCGGTCAAAC ATGACACGGT	1920
CCACTTCTTG TTCAGCTAGA AGATAGTCAA AGACAGAGTC GTATCCCTTC ATATCAGCCA	1980
AGAGTTTTTC AGACTTGACC TGAGCCAGAT AGGCTGCTGC AGCCGTATTT TGGTGCTTAC	2040
GAAGTCCCTC TGAGAAGGAA CGGAAGGATT TCTCACGAAC CTCAGCATCC TCATGGTTTT	2100
GGTAGAAATT CTCATAGGTC ACAAAGCTGT TTTGTAGGT CTGCCATGG GCTTCAAAGT	2160
CAGCCATTTC AAAATCCCCA GCTCGCATCT TAGTATAAAT GTCCTGCGGA CTGTAGAAAA	2220
CTTCACCGAG ATTTGTCAAG GCCTTCTCCA CATCTGCCCC TAAGTAGTGG GCTTTTTTGA	2280
TTTTAGCCTG ACGAATGGCA GCTGTTAAAT GTGGCAATTT ACCCAAACGG TCCAAGACTT	2340
CCTCATCTGC TGCCACCAAG GCATCGTCAA AGAAGGTCAA GGCTACGCTG GCATCTGTTT	2400
CAAATTCAT CCCAGCTTGG GCAATATTGG CAAATTCGTC ATTGCTATAG TCCGTCGTCT	2460
GAGGCATAAA ACCATAGTTG CCAATATGGC TCATCTGAAT GTAGATCTGT TCCAATCCG	2520
CAAAGGCCTT CTCGAAATCC TCAAAAGTGT GAAGATTGCC CTTGTAATCA CGGCTAAACT	2580
GGTTGATGTC TTCGCGAGCT TTCTCGATTG CACGCAAGAA ATCCTCACGG TCTTGGTATA	2640
GGGCTGTAA GTCCCAGAGT TCCTTCTCTG GAAATCTGA ACGGTGTTTT TGTTCATTT	2700
TCCTCCTCTT ATTTCTCTAA TTCTACTAAA AACTAAGGG CTGATAAAGC GTAAAGCGGT	2760
GCTGTTTCTG CTCGCAAAAT ACGAGGACCT AGGCCTGCCA AAACGGCTCC TTTAGCTTCA	2820
AAACTTTCGA TTTCTGCAGG TGAGAGACCG CCTTCTGGAC CAAAGATAAA GAGCAGTTTG	2880
GCTCCTGTTT CAAGACCAGT GACTGCTTGC AGAAGCGCAG CGGCTTCTCC TTCTTTAGCT	2940
GATTCTTCAT AGGCTACTAT GATAGAGTCA AACTGGTCCA GCTGAGCTAG AAAATCTGCT	3000
TTTTTCTCGA AAAGTTTAAT ACTTGGTACA ATATTACGCT TGCTTTGCTC GGCTGCTCCA	3060
AGGGCAATTT TTTCTAGTTT TTCAACTTTT TTACCCAATT TCTTGCCATC CCACTTGCCA	3120
ACTGACCAGT CTGCAGGAAA GGCCAGATT TGGCTAGCCC CCAGTTCGGT TACTTTTTGA	3180
CCGATGAACT CCAGCTTGTC TCCCTTGGGA AATCCAGATG CGATGGTCAC TTGACTGGT	3240

892

AGTTCACAT TGTCAATTAA TTCTTGACC AACTCAAAC GACGATTTTC CATATCCAGC	3300
ACGCGCGCCA AGCGCTTGAT GCCATCATCA AAGACTAAGG TAACCTCATC CTCTTCTTTC	3360
AAGCGCATAA CCTGAAACAT ATGCTTACTG GTTTCCTTGT CCTCGATAGT GACAGGAGAG	3420
ATAGCACTGC CTTTACAAA ATACTGCTGC ATGCTAGCCT CCAATCACAC CAGAGATATC	3480
CTTGGTTTTTC TTAAAGACAC AGGTATTCCA TTCCCCTTGA ACCATGTGAG TTTCGAGGAA	3540
AAATCCAGCT GACTCAGCCG ACTGGCGCAC CATGTCCAAC TTGTCCTTGA TAATGCCACT	3600
CATGATCAGG TAGCCTTCAT CCTTTACCAA GCGATAAGCA TCGTCTATTA GATGAATGAG	3660
GATATCCGCC AAGATATTAG CCACAATCAC ATCTGCCTCA ATTTCCACAC CCTTAAGCAA	3720
ATCTCCAGCC GCTACATGGA TATTTTCCAT GCCAGGGTTG AGCTCAATAT TTTCCTGAGC	3780
CACACGAACC GCCACATCAT CCAGGTCATA GGCGAAAATT TCTTTAGCCC CCAGAAGCGA	3840
GCTGGCAATA GAGAGAACCC CTGAACCACT CCCACATCT AGCACCGTTT CGCCACCACG	3900
AAGAACCTGT TCCAAGGCAA AAAGGCTCAT CTTGGTAGTT GGGTGGGTTT CAGTACCAAA	3960
AGCCATGCCA GGATCCAGCT TGATAATCAT TTCCCCGCA GTCGCCTCAT AGTCTGTCCA	4020
AGAGGGAACG ATGGTCAAAT CATGAGTGAT ACGAGCAGGT TCATAGTATT TCTTCCAGTT	4080
GTCTGCCCAG TCTTCCTCAG CCAAGGCAGT CGTACCTATT TTAACTCTC CCAAATCCAT	4140
AAAATCTGTC AATTCTGCTA GACGAGCCTG CAAATCCGCC TCAACCACTG TCACATCCAC	4200
CGTGTACGGG TAGTAGGCTG TCACTACGAT TTCTTCTTGC TGCTCCACCT CTGGGAAAAT	4260
CTCTCCAAAG CGGTCCACAT TTCCACATA GTCCATACTG TCTTCGATTG CGACTCCTTG	4320
CGCTCCCAGC TCAATCAAGA GATTGGAAAC CAACTCCTCT CCCTCAGGCT TCACTGTAAC	4380
TTTTAACTCT TGCCATGTTT CCATTATTAA TACCAAGCCC GTAAAACACA AAACCAAAAT	4440
AGGAAATTCT CTGAAGACGC TTGTGTCTAA GAGAAGTTTA TCTTTTGGC ACAGTGTTTA	4500
GGGCGGGTTC AGTTTAGAAA TGTAAGTAA CCATCCTTTC TAATCACTTA CTTTAAATA	4560
ATCTTTTAAT CTCTCTTGCA ACTGAGGCAC AACTTGACTG GAACTAAGAA ATTCCTCAAC	4620
ATTCATCAGC TGATAGCCCT GTCCTTCATC TCCGAAGATG ATATTGTCAA ATTGTTCTTG	4680
TCTTAGCTGA CCAACCATAA AGACCGATTT CTTGCCTTTA AAAATTACGC TAGGATAAAT	4740
CTTGCTCCAA AGCAGACAGT CTTCATCTAA ATGAATTCCC AGTTCCTCAT AAACCTCACG	4800
CCGAGCGCAT TCAAAAGGGC TTTCGTCCCC TTCACGGCCA CCACCTGGCA GTTCCCACAT	4860
ATTGGCCCAG GGAATACTTG CCTTATCATC GCGTAAGATA GTCAAAAGCT TATCCCCACA	4920
AAACAAAGCA ATCTTGCAAC CTGTGAAATC AGAAATTTCT AGTTCCATCT TCAGTTCCTT	4980
CTAACATTTT CTTTTCAGC TCGGCTAACC AGTTTTCATA ATATCTTTTC TCATCCCTCA	5040

893

ACATTCGACT ACTATCCATT TTCTGTCTAG CAATCTTGAG AGCCTTACGA GTTCGATCTA	5100
CATCTTTCTT CACCTTTAAT TGATACCAGG CTGTATCAC TTGAAGATTG GACAGTTTGA	5160
GAGACAGAAA CGATTTGACC TGTCGAATAC TAGCATATTG CTCCGCTTGC TCAAAATCTC	5220
CTTCCAACAA GGCGATATGA AGCAGGGATA GTTGGGCAAC TGTCTGCATC ATCGGAGTAG	5280
TTGTCTCTC AAGTAATGCT TGAACTGCT GTTTAGCTAC TTCTTCCTTC CCTTCCAAAA	5340
TGGAACTTC ACCTTGATA CCTAATACAC CATCCGCAAA ACTCCCTCGT GCATCCTCAG	5400
GAACTGCTTG AACAAAGTCT TTCAAATCAT ATTCCTTGAGG AGCTAGCAAG GTCTGGGCAG	5460
AATGTCTCAA TACCAGGTAG GCGTATTGG TATTTTCAGG GTGTTGTAGT AATTCCCAAA	5520
TTTTTGCTCC ATCGGTGATG TCGACTGGCA AAATGTTATT TAGGAAGAAA GATAAATTAA	5580
GAAAAATCCA AGTCCCTGCA AAATACCAGC TTCTTGTCAA AAATCCAAAC AATATCGCCA	5640
ATAATATCAA GCCGAGATGA ACCATCAAGC CTCCTGAAAG CATCAGGATG ATTCCTTTGAT	5700
CGCTTTCATC CTCTTTTAAA CCAATGTATT GAGCACCAAC ATTTTTCAGA ATGGCTGTTT	5760
TACTAAGATG AAACCTGCCT GACTTTTTGG TCAAAATAAA ATGTCCTAAT CCAAAAGCCA	5820
CCAGCCGATA GCCTGTCAAG TAGCCACAAA AAGCATGACC CAGCTCATGA AGAATAAAGA	5880
TTAAATACAT GCTTAGAAGA GCGAAGGCAT AACCAAAAGT AAAGGCTAAA ACTGCGGAAT	5940
ACCCCAACTC TGCAAATGCG ATTGTTCCAC AAGCAAAAGC TAGCATAATA AAGACAACAG	6000
CTAGCACATA AACCAAATAA GTCCCAATTT TCTTCATAAC ACCTCCAACC AACTCCTAGT	6060
ATCTTGATA AGGATAAAAT TCTCCCTTTT CCAAGCCAAT TTTTCCTTCT TCAAAGACTT	6120
CTTGTTTCCA TTCCATGACA AATTCCTCTG CTCTGGGTC TTCCAAAAAG TCCATGAGGA	6180
CATCTAGCCC AACCTCAGCA GTATCTTTAA GGAAAAGCGC AAAATAAGCT AAAAATTCAC	6240
GGGAAAATCC TTTTITAGGC AGGTAAGGAA TAACAGTCAA ATAGTCTTCC TCATTGACTG	6300
TTGACTTGGC AGGATTGTAG AAAAGGACCG CTTCCTCAAA AAGAATGTCA TCTGATGAAA	6360
CCTCTCCGTC TTCATCCACC ATCTCCACAC CGCAGCATTT TCGCTTCCA ATAGAAAACT	6420
CACTTCTACC GCATGGTGC GTTGTGCCA GCTAATCTCA AAGTCAAAGG GAAAGTTCTT	6480
GTCCAACCTT TCCTCTAAAA TATCTAAAAA TCCGTATGTT GCCATTTTGT CCTCTTTCTA	6540
TGCGACTCTT TAATCGCCCC GATTGCTCGG AAATATGCTA AAATAGATAC TACCATCTTA	6600
CCACAAAATT ATTTTATGTC CTAATTATAC CATATTACCT CATTTAAACC CTGTTATCA	6660
GTGATTTTCT TAAAAGTCTG ATTTCTTCAT TTCTCATAAA AATCAATATA AAAAGCCCTC	6720
GAAAGGGCTA ATAAATCTAT AAAATCAATA GGCGAGTAAC TAGCACAAGT GGACGTGCTT	6780

894

TTTTATTGAC TATTACCACG ATACCACGCT TAATCTTAGG CTTGAACTTT CTTATCTGCA	6840
ATAGCGTCTG TCAAAGTCTG AGAAAAGTTA AGCCCCATTT CTCGTCCCAA CTTATCTGCC	6900
CATTTTGGA TGGTCAAAGT CTTTTTAATG GGTTCCTGAC TTCCTAGGTA TTCTGATACA	6960
TCAACAGATA CCATAGAAAT AAAAGATTTA TCAAGGTCAT AGGTTGACAC GAAATCTTCA	7020
TCATCTTTAA AAGGATCATT ATCAATTAAA GACAAGCTAT TGATATCTGA TGGCTGAGGT	7080
AACTCTCCAT CACTCTCTAT CAAATCTGCA ACAGTTATCC CTAGCCACTC CGACCCCAT	7140
GCCAAAGCCT CAGAAATCCC CTCTCCTTGT GTAGCTGAGT ATTCAAAATC TGGGAAATGG	7200
ACAAAATAAG TCGCTTCTGT TCCGTCTGTG TCGTCATAAT AAAATAAAGC TGGATACGTA	7260
ACTAACATTT CACTACCTCC ATATCAAAAA GCAGGGACTG AATTTTACAA CCCAGCTTGC	7320
TTTCTTATCC CTCTTTCAGT GTACTTATTC AGCTCACCAT GAAGGATTGT GATAGGTCTT	7380
TCCCCTTGCT TTTCCATTTT AATATGGGAG CCTTTACCGC CTCTAGTCTT TATCCAACCA	7440
TGGGCCGTAA GGAGTTTAAC CATCTCTTTT TGTGTCATAG GCATAGCGCT TTTACCTCCT	7500
GACAACACCA TTATAACACG TGTTACACGT ATTGTAAAGG AGTGATACTT ATTATTCTAT	7560
TATACATAAA AGCCCCTAGA TGTGGTTCTA AGGGAAGCCA ATTTATTCTAT ACCTATTTTT	7620
CTAATGAGTA GTAAAACTG CTTCTTTATC GAGCAATTCA TCATCTGTAT AGTCAATTGT	7680
AAAAGTATCT CGATCTAAGA CAGATTGAGG CGGAGTTGAA TGAATCATAG GAACACTGCG	7740
TACTCTATAT TTTTATCTC CAATTTTAC AAACGTATAC TCTTCGAAAA TCAAATTCAA	7800
ACCACGTCAA CGTCGCCTTA CCGTACTCAA GTACAGCCTG CGGCTAGTTT CCTAGTTTGC	7860
TCTTTGATTT TCATTGAGTA TGATTAACTC TCAAGTCTTC GAAATCAGGA TTTTCAACAG	7920
TTATTACAAG GAGGCGATTT ACTACTTCAA AAACATCAAT TATCTATTT TTCATATTTT	7980
TTCAACCCAT TATTAGAATG AACTTCTTGG TAAGCAAAAT CAAGTTTAGA TTTAATGTTT	8040
TCGTACAAAT CTAAAATCTC TTTTGGAGTA TCTTCCCGGA AGAAAAGTTT TCTTTCCCT	8100
GAAATAACTT GATCACTAAG AATCCAATGA CGAATTTGTT TTGTAAAAAT CAAAATTGCC	8160
TGACTTGGA GTTCCATCAT TTCCATTGCT TATCACCTCT CTTTTCATTA TAGTTCATAC	8220
AATGACATTC AGCAATATTA TTTCTCAAGT CAGCACTTCC ACTTCTTTAG GCTCAACTAT	8280
CCTATTTTGA GCTTTAAGGA AAATCAAATC TCTCATGCTG ATACCTCTCC TCATTAAATT	8340
AAATAGTAAA AAAGATTCTA TCTCACTCCC TGATTATTAC AAAACCATTG AAATATCACA	8400
ACTAATAGGC TAGAATGGAC ATAGTAAGAT ATAGTAGATG AGTCATTCTA CTCAAATCCA	8460
CGTTAGAAAG GACTGCTATG CCAGACAATC TCGCCGTTGG CATGCGCCCG GG	8512

(2) INFORMATION FOR SEQ ID NO: 130:

895

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2869 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

CTCGTTTCAA GGTGAGTCT CTTGCAAATC TTGTTGCGGT TCTTCCTTTT GCCAAGGCAT	60
CTCTCCCATG GTTGGTGCCA GCCATTGTTG GAATCTTGCT CTCATTGGTT CTACCAAACA	120
AGCAAGAAAG CGATGTTTTT GAAATGGAAT AATCACTTAA ATCACTTTTG TAGCCAAGTC	180
TACAGGAGTG ATTkTCTTTT TTTATCCGAT GATAAATGTG TTATAATAGG TAGCGAAAGA	240
GGTGAAGAAA TGAATCAAAC AGTAGAATAT ATCAAAGAAC TGACAGCCAT TGGCtCGCCA	300
ACAGGCTTTA CTCGTGAGAT TCGGACTAT TTAGTCAAGA CTCTAGAAGG TTTTGGTTAC	360
CAGCCGGTTC GCACATCCAA GGGCGGTGTC AATGTAACTA TTAAAGGTCA AAATGATGAG	420
CAACATCGCT ATGTGACTGC CCATGTAGAT ACGCTTGGTG CTATTGTCCG TGCTGTCAAA	480
CCAGACGGCC GTCTCAAAT GGACCGTATC GGTGGCTTTC CTTGGAACAT GATTGAAGGA	540
GAAAACGTGA CCATTCATGT GGCTAGCACA GGTGAAAAAG TATCAGGAAC CATCCTCATC	600
CACCAAACCTT CTGCCATGT CTATAAGGAT GCAGGAACTG CAGAACGCAC GCAAGACAAT	660
ATGGAAGTGC GTTTGGACGC CAAAGTAACT AGTGAAAAAG AAACCTCGTC TCTTGGCATT	720
GAGGTCGGTG ATTTTATCAG TTTTGACCCA CGAACTGTCTG TGACAGAGAC AGGTTTTATC	780
AAGTCTCGCC ATTTGGATGA CAAGGTCAGT GCGGCGATTT TGCTCAATCT CCTTCGCATT	840
TATAAGGAAG AGAAGATTGA ATTGCCCGTA ACAACTCATT TTGCTTTTTT AGTCTTTGAA	900
GAAGTGGGAC ACGGTGCAAA CTCTAACATT CCTGCTCAGG TAGTAGAATA TCTGGCTGTG	960
GATATGGGAG CCATGGGAGA TGACCAGCAA ACAGACGAAT ATACAGTGTC TATCTGTGTC	1020
AAGGATGCTT CTGGACCTTA TCACTATGAC TTCCGTCAAC ATTTGGTGGC TTTGGCGAAA	1080
GAGCAAGATA TTCCATTTAA GCTGGATATC TATCCATTTT ATGGTTCGGA CGCTTCAGCG	1140
GCTATGTCTG CAGGGGCAGA AGTCAAACAC GCCCTTCTCG GTGCTGGTAT AGACTCTAGC	1200
CATTCTATG AGCGTACCCA TATTGACTCG GTGATCGCAA CAGAACGAAT GGTGATGCT	1260
TATCTTAAGA GCACGTTGGT GGACTAATAT GTGCCTTATT TGTCAGAGAA TTGACCTCAT	1320
CAAGAAGGAA GAAAATCCTT ACTTTGTCAA AGAGTTGGAA ACAGGCTATC TTGTGGTTGG	1380
AGACCACCAG TATTTTGAAG GCTATAGTCT CTTTCTAGCC AAGGAGCATG TCAGCGAATT	1440

896

GCACCATTG AAAAAGGAGA CAAGACTCCG TTTTCTAGAA GAAATGAGTT TAGTCCAAGA	1500
GGCAGTTGCC AAGGCCTTTG CTGCTGAGAA AATGAATATC GAACTGCTAG GAAATGGCGA	1560
TGCTCATCTT CATTGGCATC TGTTCACAG ACGGACAGGT GATATGAATG GTCATGGTCT	1620
CAAGGGTCGT GGACCAGTCT GGTGGGTTC CTTTGAAGAA ATGACAGCAG AAACCTGCCA	1680
AGCAAAACCG GATGAGATTA AAAGATTAGT CAAACGTTTA TCGTCAGAAG TAGATAAACT	1740
ATTAGAAATA AAGGAGTAGA AATGAAGAAA AGATACCTAG TCTTGACAGC TTTGCTAGCC	1800
TTGAGTCTAG CAGCTTGTTT ACAAGAAAAA ACAAAAAATG AAGATGGAGA AACTAAGACA	1860
GAACAGACAG CCAAAGCTGA TGGAACAGTC GGTAGTAAGT CTCAAGGAGC TGCCCAAGAG	1920
AAAGCAGAAG TGGTCAATAA AGGTGATTAC TACAGCATTC AAGGGAAATA CGATGAAATC	1980
ATCGTAGCCA ACAAACACTA TCCATTGTCT AAAGACTATA ATCCAGGGGA AAATCCAACA	2040
GCCAAGGCAG AGTTGGTCAA ACTCATCAA GCGATGCAAG AGGCAGGTTT CCCTATTAGT	2100
GATCATTACA GTGGTTTTAG AAGTTATGAA ACTCAGACCA AGCTCTATCA AGATTATGTC	2160
AACCAAGATG GAAAGGCAGC AGCTGACCGT TACTCTGCCC GTCCTGGCTA TAGCGAACAC	2220
CAGACAGGCT TGGCCTTTGA TGTGATTGGG ACTGATGGTG ATTTGGTGAC AGAAGAAAAA	2280
GCAGCCCAAT GGCTCTTGGA TCATGCAGCT GATTATGGCT TTGTTGTCCG TTATCTCAA	2340
GGCAAGGAAA AGGAAACAGG CTATATGGCT GAAGAATGGC ACCTGCGTTA TGTAGGAAAA	2400
GAAGCTAAAG AAATTGCTGC AAGTGGTCTC AGTTTGAAG AATACTATGG CTTTGAAGGC	2460
GGAGACTACG TCGATTAATA CTCTTCGAAA ATCTCTTCAA ACCACGTCAG CGTCGCCTTA	2520
CCTACTGACT GCGTCGGTTC TATTCACAAC CTCAAAACAG TGTTTTGAGT CGATTGCTCA	2580
GTTTTATCTG CAACCTCAA GCTGTACTTT GAGCAAGCG GCTAGCTTCC TAGTTTGCTC	2640
TTTGATTTTC ATTGAGTACA AAAAGTAAAC TTTTCTCTTG CAATTCCAGA TAAATAGTGT	2700
ATAATGGATG GGTATGTGAA AAACATACTT GTGGGAGGTA AAAATCTCTA ATTACCGCCA	2760
AAACCACAAA GGAGGATTTA AAAATGGCTA AAAAAGTCGA AAAACTTGTA AAATTGCAAA	2820
TCCCTGCTGG TAAAGCTACA CCAGCTCCAC CGGTGGACC TGCTCTTGG	2869

## (2) INFORMATION FOR SEQ ID NO: 131:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6186 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:



897

CTGAATCCCT TATAGGAGTC CAGTAACTTT TTAGCCTCTA CTTTGCCTTC ATAGGCAGCT	60
TCAACATCAT TAAAAAAGA AAGCACTGAA GCAAGTTCTT CAGTGCTCCA CGACAAATCT	120
AGTGGGTAAC TATACTGTTT GTTCATTAAC TAATACCAGC TCTCATTCTT GCTTCTTTTA	180
GTTCCTGCTT ACGATAACTA CGAGGGAGAA AAGCACGAAT CTCATCTTCA TTAACCCGA	240
TTTGCATACG CTTGGCATCA ATAATAATTG GACGACGCAA AAGACTAGGA TACTGCTCAA	300
TCAAATGAAG CAATTCCGAT ACCGAAATAC TCTCTACATC AATATTCAAT TTTTGAAAAA	360
TTTTTGAACG AGTTGAAATG ATGTCATCAG TACCATTTTC GGTCAAGGAA AGGATGTGTT	420
GCAATTCTTT TCTGTGTTAA GGACTGGTCA TAATATTGTG TTCCACAAAG GGAACCTATG	480
TTTTTCTAAC CAGGCCTTAG CCTTACGACA TGATGTACAG CTCGGTGATA GAAATAGTGT	540
AATCATGCTT TTCTCTCTT ATCTATACTT TGCTACTTCT ATTATACAAA AAAATAAAGC	600
GCTTGACTAG GGATTTTTAG AAAAAAGCC TATTTTTTCA AGAAAAATAG GCTTTTTGCG	660
AACGATTGAC ACAATTGGAT TTGGTTAATT CACTCTTAAC GATGGTTTTA AACGATATAT	720
ATTTTTATAT ATGTAAATTA AAAACATCTT TCCTTTCACT TCCTACGACT TTTGAGATAC	780
AGATAGCCAA AGAAGTTTC ATAGAGGGCA AAAAAAGGA GGAAGGCATG AAGAAAGAAG	840
GTCTCTGGCA AAATCATAAT AACAGGATCC TTGGCTGGAT CAAAAAGCCA GGTATCATCT	900
CCCACAAAGA GAATTTGATG GAAAAGAGTA AAGAATTGGT CAAAACCAAT CAAAACCTCC	960
CCAAGTCCAA TCATCACAGG TAAGACTACT AGAGCCAGGA GACTTTTTCG ATAAAGAGAC	1020
AAAAAGTCCT TTTTACAAT CCTATTGACA AAGACATAGA AACTTGGCAG TGCTACTAGA	1080
GCTACTAGCT GAACCAAATG AAAGAGATTC TTGACCACTG CGAAATGGTG CAGACCAGCT	1140
GCTGACGAAC GAAAATCAGG CATCTGTAAG ACCTGACTAA AAGGATTGGT CAGATAATTC	1200
ATCAAGATAT GAAAATTGTA TTGAATGGTT TCTGGTTTTA GATAGACTCG ATTCGTTAAG	1260
TTTAGCCACT GAATCTCCAT AGGATAGAAA ATCCAAGCCA GATAAATGGT CAGAAGGATG	1320
GAGAGGGAGA GGAGAAAGAG CATAGAGCCC CAAAAGATCA ATTTAGTTTT CATCAAAATC	1380
CCACTCCGCA AGGCTAGAAA CCACATGTGT CGGTGCGATT GGCAGGCCAG CTACTTCTTC	1440
TGCCTTAGTA AAACCTGTCG TCACCAAGAG CGTTGGAATG CCATTGTCAA TCCCAGCCCC	1500
AATATCAGTC AAATAATTGT CCCCACCAT GATTAACCTC TCACCTTCCA AACCTAAGTG	1560
CTCAACCGCC TTGTCCATAA TGATGGCATT TGGTTTTCCG ATATAAACCG GCTTCACTCG	1620
TGTCGCTACT TCAAGCAGCG TAATCAGTGA GCCAGCACCT GGCAAAAGAC CGCGTTCCGT	1680
CGGGATGTTG AGGTCAGGAT TGGTTCCGAT AAAATGGGCA CCCTTTTGAA TAGCAAGAGT	1740

898

TGCTGTGGCA AATTTTTCAT AGTCGACTTG CCAATCCAGA CCAACTACCA CGTAGGCAGG	1800
TTTTTCCTTG TCTTCCACAT AACCAGCCGC CTTGATGGCT TCCTTGAGTC CTGCTTCTCC	1860
GACGACATAG ACGGTCTTTT CAAGCCCCAA ATCATTCATA TAGTCGATGG TTGCCAAAGT	1920
CGCTGTGTAG ACAGTCGATA GGGGCGTATC GATATTAAAA TTCTGAGCCA ACATCTCCTT	1980
AACACTCTCT GGAGTGC GGG TTGTATTGTT GGTTACAAAG AGATAGGGAA TGTCCCGCTT	2040
TTGCAATTCA TGAACAAAAG TCTCTCCAGC AGGGATTCCG TCTTTCCCTT TATAAATGGT	2100
TCCGCTCTAAA TCAATTAAAT AGCCTTTATA TTTTCATCTAT TTCTCCCTAA GCCTTTTTTA	2160
TTTCTTGCCA AGTAATGATT GCTTGGGCAT TGATAACCCC ATCACTTGTA ATTTTCATGCT	2220
TGCTTCCAG TCCAGTCCGT TCAACAGCCG ATGTAATCAC CCCACCTGGT CGAACTTCCT	2280
TGACATACTT GAGGTTGATT TTCTTGGGAA TATAGTGGGT CAAAAATCC GCTCCCATGA	2340
CCTCAAAAAT CCAGTCCAAG TATTTACTGT TATTGACATG ACCATTCATA TCCAAGTCGT	2400
AAAAACGAAC ATGGTAATCC TTGCTGATCG GTTCTTCCAA GGA CTCATAC TTCGGTCCAC	2460
GGATAAGTTT TTTATCAAAA TCAGACTGGT AAGGAGCCAC AATCTCAGGT TCAACAACAT	2520
GGACTTTTCG ACTGTCCGGG TCCATGAGAA CAAAGGTCCG CATCATGTGG ATGAGCTCCT	2580
GCTCCGCTTC ATTATAAATA GTAAAGCGAC GGTAGCAAAA AAGTCGATTG TAGCTCAAGG	2640
CTTCCGTTTC GATGGTAATT TCTTCCGCAA AACGAGGCAA ACGAACCACC TCAATATCAT	2700
ATTCTACGAT AATCCAGACC AGATTATATT CTTCCAAAAT GGCCTTATCA CTAACCTCCA	2760
GTTCAATCGA CTGCATCCCT GAAACTTGCA GTGACAGCAA AATCACATCT GGAAGTTTGA	2820
TATGACCGTT CATATCAGCC ATATCAAAAAG GAATTTTCAT TTTTCATTTGA TAAGTTAAGC	2880
CCATGATCCT ACTCCAAAAT AAATCGTTCT GCTACAGTAT CTCCCAAAAA GAGACCTCTC	2940
TTTGTCTATG GAACGTGGTC ACCCTCAATC TGCATGAGGC CTTGTTGAAC CAAATCTCTG	3000
ACAATTTCTC CATAAAGTCC AGCAAAAGAC TGTCCAAATT TTTCTCATA TCGCGCCATG	3060
GAAACCCCGG ATTTCTTGCG GAGTCCCAAG AACATTTCTT CTTCCATTTG CTCCTTTTGA	3120
CTCAGGTGAT CTTCTGTAAT ACAAGCATTG CCTTCCTCAA CCGCACTGAG ATAATGACGA	3180
ATGGGACCAT GATTTTATA GCGTACTCCA TTGACATAAC CAGATGCCCC TGCACCAATA	3240
CCATAGTATT CAGCATTTG CAGTACATG AGATTATGAC GACTTTCAA ACCGGGTTTG	3300
GAGAAATTAG AAATCTCATA ATGCTCAAAA CCCGCTCGCT CCAGCTCTGC AATGATGTAC	3360
TCAAACATCT CCCCTTCTAG TTCCTCCTTA GGCAGAGGCA ATTTCCCACG TCGCATCCGG	3420
TTCATAAAGA CCGTATGGTT TTCTAAAATC AAATATACA AACTCATGTG GGGAAATATCC	3480
AATCCAATGG CTTTAGCCAC ATTTTCCTTT ACTTGCTCCA TGGTCTGACC AGGCAGAGCA	3540

TAAATCAAAT CAATGGAGAT ATTGTCAAAA CCAGCCAGTT TCAGGCGATC GATATTTTCA	3600
TAAATATCCT TCTCCAAATG ACTGCGCCCA ATCTTTTTTCA ACATCTTATC ATCAAAGGTC	3660
TGGACACCTA GCGAAACACG ATTGACAGCC GAATTTTTTCA AAACAGCTAT CTTATCCGCA	3720
TCCAAATCGC CTGGATTGGC TTCAATGGTC AACTCTTCCA AGACAGACAA ATCCAAGTTT	3780
TTAGTCAAGC CATTCAGTAA CACCTCCAGT TCGGAGCCG ACAGGGCTGT CGGTGTTCCA	3840
CCACCGATAT AAAGGGTTGA CAACTTTTCA ATATCATAAG AACGAAACTC TTCCAGCAGA	3900
TGCTCTAAAT AGCTGTCGAC TGGCTGATTT TTGATGAAGA CCTTTGAAAA ATCACAATAA	3960
TAACAAATCT GGGTACAAAA TGGGATGTGC ACATAGGCTG ACGTTGGTTT TTTCTGCATA	4020
GTAATTATTA TACCACAAAG ACTAGATTCC AGATAAAAAAT CACCATCCCC AGATACATAG	4080
TCCGTCCGGA GATGGTGATG GTTTATTCTT CTGTTATATC AATCACAATC TCTTCTGAGT	4140
CATCAAGAGC TTCGGCTTTT TCTTGCCATT GCTCCTTGAG ATTATTTAAT TGATTTTTTG	4200
ATGCTTCTGT CGCTTGAAAA GCATAGGATT TAGTTTGAGC AAGTATACTG TCCACAGTGA	4260
TTTACCTGA CTCAACCTGT TCTTTTGTTT TCAGAACAAA ATCTGTAGCC TGCTCCTTAA	4320
CTTCTGTCAG TTTTTCACAG ACTTGCTCCT TGGCATACTC CGGATCTTCT CTCAAATCAT	4380
CTAGAAAATC TTGAGCCTGA CTGCAAACCT GTTTGCCCTT ATCACTTGTT AAAAAACAAGG	4440
CAAGAGCTGC ACCTGAAACG GTTCCTAAAA GGATTGAGGA TAATTTACCC ATAAGGATTC	4500
TCCTTTTTTA TTTTTGAAA AATTTACTTG CAAGACGAAG AGCTGACAGA CTGCAACGAG	4560
TCTTGAGTGT TTTTGAACCA GCTGATGAAG CTTTCTTGCT CAAGACACGC GCATGGTCAT	4620
TGAGGTCTGA AACAGATAGA GATAAATCTG CAACAGCACT GAAGAGTGGA TCAATCGTAG	4680
CCACCTTGAC ATTGATATCA TCTGCCAAGA CATTGACCTT AGCCAACAAC TCATTGGTGT	4740
GATGCAAGGT CACATCCACA TCTGAAGTCA AGGTTTAAAT CGTCTTTTCT GTTTCATCGA	4800
TGACACGACC AAGCTTTTGT ACAGTAATGA TCAGATAGAC CAAAAAGACA ATCAAAGCTA	4860
GGGCAACAAG AATATATGCA ACTTCTAACA TTTAGTTTTT CTCCTCTGTA ATATAGTAAG	4920
GGGCCTTCTT TCGATTTTGA TAAATAACGA TCATTATACC GAGACCGATA AGGACAACCTG	4980
ACAGCCATTG GGACACTCGA AAGCCGAAGA ACATGAGACT ATCTGTTCCG ATACCTTCGA	5040
TAACCATACG ACCGAAACCA TACCAATCA AGTAAAAGGC CGTGATATGA CCTCGTCTGA	5100
GACTCTTCCA TTTCCGTCTA AAAATCAGAA TCAAGGCAAA GCCAAGCAGA TTCCATAGAG	5160
ACTCATAAAG GAAAGTCGGT TGACGGTAGC TCCCCTCAAT ATACATCTGG TCACGGATAA	5220
AGCCAGGTAG ATAATCCAGA TTATCCACTG TTGCACCATA AGCTTCTTGG TTAAGAAAT	5280

900

TACCCCAACG CCCCAAACCTT TGAGCAATCA TAACGCTAGG CGCCGCAATA TCTAGAAAAT	5340
CCCAAGTATT GATGAGTTTA CGGTCAGCAA AGATATAGAG CACAAGAGCC CCAGTTATCA	5400
AACCACCGTA AATGGCCAAA CCACCATTCC AAATGGCAAA AATCTCTCCT AAATTCTGAC	5460
TATAGTAATC AAATCGGAAA ATAACATAGT AGAGACGAGC TCCTAAAATA GCCAAGGGAA	5520
AGGCTACTAA GATAAAATCT AAAATATCGT CTGGTATGAT CTTCTTTCTA GGTGCTTCTT	5580
TCATGGTCAA ATAAACCGCA AGAATCAAGC CTGTCACAAT ACATAAGGCA TACCAACGAA	5640
TGGCTAGGGG TCCTAGTTGA ATAGCAATTG GATCAAGCAT TTTGCACCTC ATTTGAGCG	5700
ATTAGACTTG TCAGTCGTTT GTCGAACAAA CGGGTCGCAT CAAAGCCCAT TTCCTTGGCA	5760
CGATAATTCA TGGCAGCTGC CTCAATCACA ACAGAGATAT TACGACCTGT TTTAACTGGA	5820
ATACGAATAC GAGGAATGTA CGCCAGAAAC TTCAAGTTCC TCTGCATTAT TTCCAAGACG	5880
ATCAAAGGTC TTATGCGTAT CGTAATTTTC CAAATAGACA GCAAGCTGAA CCTGTGAAGA	5940
ATCCTTGACA GCACTCGCAC CGTAGAGACT CATAACATCG ATAATACCA CCCCACGAAT	6000
TTCAATCAAG TGTTCAAAA TTTCAGCTGG TTCACCCAG AGAGTAATCT CATCCTTGGC	6060
AAAGATATCG ACACGGTCAT CGGCTACCA ACGGTGACCA CGTTTGACAA GCTCAAGACC	6120
TGTCTCGCTC TTACCAATTC CACTATCTCC CTGAATCAAG ACGCCCATCC CATAAATATC	6180
CATCAA	6186

## (2) INFORMATION FOR SEQ ID NO: 132:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9541 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

GAAAATCACA ACCCTTTTTG CAAAATTTT GAGATTATTT TCACAACTT GATTTTTC	60
AGTATACTCA ATAAAAATTA AAAAAATCCA CTACGTCAAG GCGAGGCTAA TGTGGTTTGA	120
AGAAATTTTC GAAGAGCGTG AATGAGTATC ATCTATAGTA AAATAAAAAA ACTGAACAAT	180
TTGGTTGGGG ACAGCCAAAC CAATTTCTCA CAATGTTTCA GAAACAAGGG TGTGCTATTC	240
CAATTTCAGC CTAATAAAC TGTATAGAT TGCTGAAACA AAGCTAGGT AAAAGTCTTC	300
ATAATAAAAA GACCTCCTAT CAAGTGTTCA AAACTTTGA TAGGAGGTCT TGTTTTGTGA	360
AAATATTTAT CAAATTTTCT ATACAAGTGA GCTGTTAGCC AGGTTCTTTC TATTCTTTCA	420
ATTTCAATGA ATGGATTTT TACTAATACT CATAACTGGG AATTTGCTG TGTAATAATA	480

901

GCGAGATAGA TGGTATTTAT AAAACACTCA AGACAGCTAG ACTAATATCA TTTAAAACAT	540
TATCTTCTTT TGAGCGACTG TTGGTTACCA ACATAGCTAA ATTCCTGCA TTTTCAAATT	600
GATAGGGTTC TGATTTAGCA TTCACAACCA CCAAGAGGTG TTCTTTGCCG TGAAC TTCAT	660
AGATAAGGTA GCCGCTATGT TCAATCGCAG AATGCACAAA GACATGATGG TAAAT TTCAT	720
CATAGCTAGA GTAAGAAAAG GCACCAGTTT TTGTCTTCAA TCGGATGACT TGACGGATAA	780
ACTCAATACT GTCTTGACGC TCATTAATCA AGTTCCAGTT CACTTGGTTC AACTGTTCAG	840
GAGCATTATA GCTATTCATC GCACGCTCTC TATCATCATG GGTCAACTCA CCATTTTCAC	900
CAGTCGCAAC CAGTTTGGTA CGACCAAATT CTTGACCGAT TTCCATAAAG GCCATCCCCT	960
GCATGAGCAG ATTTCATGGCT GTGGCAGTTT CGACCTTGGC CATGATTGTC TCTGAACTTT	1020
GGTCTGGATG AAGGGTTGCC AATAAATCGT GAAGATTGTA ATTGTCATGG GCTTCTACAT	1080
AGTTAAGCAC CTGATTGGGA TGTGTATAGC TTCCTAATTC ACGACTTCCT AGGATTGCTT	1140
TAGCTAGAAT TGGCTCTGTC GCAGCACCAC TGACAAAACC TGACTTGATA GCACCATAAA	1200
CTTCTCCCCC TTTGACAGCA TCGCGCTGAT TGTCAATAAA GAAACCAATA TTTGGCATCT	1260
GGTAGGCATT GTCCTTCTTG GCCTTATCAT AAGGGGCAAG ACCTGTTCCC ATATCCCATC	1320
CTTCTCCATA GAGGATAATG TTGGAGTCGA TTTCATCCAA GCTTTGACGA ATCATCTGCA	1380
TGGTCTTGAC ATCATGAATC CCCATCAAGT CAAAACGGAA GCCGTCAATA TTATATTCCT	1440
GCACCCAGTA TAGAAGAGAA TCAATCATAT ACTTGCGAAA CATTTTCGTGT TCACTGGCTG	1500
TTTCATTTCC AACACCCGTT CCATTCTGGA AGGTACCATC TGGATTGATA CGATAATAGT	1560
AATCAGGGAC TGTGTTTGG AATGGTGCA CAACAACCTGA GAAGGTATGG TTATAGACTA	1620
CATCCATAAT GACTCCAATA CCCGCATCGT GATAAGCTTG AACCATCACC TTCAAATCAC	1680
GAATGACCTG AGCTGGATCA TCTGGATTAG TTGAAAACT AGTTTCTGGC GCGTTATAGT	1740
TTTGTGGATC ATAACCCAG TTGTAGGTTA CATTTCCATC CTCATCGTAT TCTTTATGAC	1800
GGTCTGCAAT TGGTTGCAAT TGAACATAAT TGTAGCCCAG CTTCCTTGATG TAATCAAAAG	1860
CAGTTGACTG GCCGTATTGG TTAACGTTC CAGCCTGAGC AGCACCCAAG AAAGTTCCTC	1920
GAAGATGTTT ATCTACACCC GATGTAGGTG ATTTAGTCAA ATCACGAATG TGCATTTTAC	1980
AGATAACTGC CTTACATGGA TTTTCCAAGC GCCAAGTAGC CTCCGAACCG TGCTTAACCT	2040
CGAAGTTTTC AACTTGCTTT TCTACATGGC TCAGAATAGC TGAACGTTTG CCATCAGGGC	2100
TGGTCGGAT TGTATAAGGA TCACGTGTCA GTGTTTGCTG ATGAGGGAAT TGGACTTGAT	2160
ACTGATAAGT CTTACCTACC AAATCTTCTT CAACATCCAA ACTCCAGACA CCGATTGTAT	2220

902

TGTCCTTATG ATTATAAGAG TAGCTATTGC CTCTTTTCAT CTCAAAAGTC TTCCAAACGG	2280
GTGCATCATT AGCAGCTGAT TCATAAACGA CAACTTGCAC TTCTGTGCT GTAGGTGACC	2340
AGAGAGAAAA ATGAGCCTGA TTGTCCTCTA CACGGCAACC CAATTCTCCT TGGTAACCCC	2400
AATGATGATC AAAACTAGCA CTGTTAATGG CCTTATCAAA GGCAAAAGGA TTTTGATTTT	2460
TATAGAAAGG ACTGGCAATA GCAGGATTTT CAGAGTAATA AATCCTATCA TCGCCTTCCA	2520
AAATCCAGAC CTCTGTTAAT AGGGGATAGT GATTAAAACG GATAGAATAT TCTTTACTAG	2580
TTTGACCTGT ATGAACCACA AAATTCAAGC TTTCTATAAC ATGTGAACTT GGGTGTTCAA	2640
AGCTAAATAA AGCTCCAAAA TAATCTTCTT TGTAGGTTAG CAAATCAATT CGTTGATCCT	2700
GACTTTTAC AAAGGAGCAA GTGTCATATT CTCCATTCTT ACGATGGTAA TGAATGCGCA	2760
TAGGGTAGTT ATACATTTTT TATTTTCTT TTTTACTTTG TTTCTATTC ACTAATAAAT	2820
TTTTGTCAAT CTCGTCTCAA TTAACAGACA TAGTCATATT CTCTAAACTC TGTTTTTAAA	2880
CGATCCATTA CAAACTTTCT AGCCATGCCT CATCTCTGAC CTGGATACCA AGTTCTTGTG	2940
CTTTTGCAG TTTACTTCCA GCGTCTGCAC CTACCACGAC GAGGTCGGTC TTTTAGAAA	3000
TACTACCTGT CACTTTGGCA CCCAGACTTT CGAGTTTACT TTTAGCTTCT GAGCGCTTGA	3060
GTCGTTCCAA TTTTCCTGTC AATACCACGG TCAAACCTGA CAAGGCCGCA TCCGCTACTA	3120
CCGTCTGTCC TTTATAGTCC AGATTGACCC CAGTTTCTTT CAATTCTCTG AGCAGAATTT	3180
CAGAGCCTTC TGTCGCAAAA TAAGTCTGAA GACTTTTGGC AATCACGCCA CCTAGACTTT	3240
CAATACTAGC CACTTCCTCT GAATCTGCCT GAGACAGATT TTCAATTGAA TGGAAATATT	3300
GAAGTAAAAG CTGACTAACC TTGCTTCCGA CATGACGAAT TCCCAAACCA AATAAGAGCT	3360
TCTCGGCAGA ATTTTCCTTT GATGCTTGA TAGCCTGATA CAGTTTAGCA GCGGACTTTT	3420
CCTTAACTCC CTCTAAAAGG AGGAAATCCT CTTCTTGCAA ACGATAAATA TCCGCCACAT	3480
CCTTGACTAA ATTAGCAGCA AAAAGCTTCT CAACAATAGA TGGACCAAGG CCTGTAATAT	3540
TCATAGCATC ACGAGAAGCA AAGTGAATCA AGCCTTCCAT GATTGAGCA GGGCAACGCG	3600
GATTGATACA ACGTAGGGCC ACTTCATCTT CAAAGTGCAA CAAGTCAGAG TTACAACCTG	3660
GACAGTTTGT AGGGATATCT AGTTTTTCTT CAGAAACCCG TTTGGACTCT ACCACACGTA	3720
AAACGGCAGG GATGATGTCA CCAGCCTTAT ATACAATGAC CGTATCGTCT TTTCGGATAT	3780
CTTTTTCAGC AATATAATCT ACATTGTGCA GGGTCGCACG GCTAACAGTC GTACCGGCAA	3840
GTTGTACTGG TGTTAGATTA GCAGTTGGAG TTACAACACC GGTACGGCCA ACTGTCCAGT	3900
CAACTGATAA GAGTTGAGCT TCTTTTCTT CGGCAGGGAA CTTGTAGGCT ACTGCCACT	3960
TTGGAGCCTT AACTGTAAAA CCAAGTTCTT CTTGACTTGC TAGGTCGTTG ACCTTGATTA	4020

903

CCACTCCATC AATATCGTAA GGCAGATTTT CCCGTTCCCTG TCCTACTTCT TGGATAAAAT	4080
TCCAGATTTT ATCTATGTTT TCAGCCAAGA TTCGCTTAGG ATTGACCACA AAACCTAGTT	4140
GTTCTAGGTA CTTCAAACCC TTTTCTTGGC TATCACGAGT TGAAGGGCTG GCTTCTTGAT	4200
AGAGAAACGT TGCAAGATTA CGCTTGGCAA CTACTGCTGT ATCCAACGA CGCAGAGTTC	4260
CTGCTGCCGC ATTACGAGGA TTAGCAAATT CAGGCTCTCC ATTTTCTTGG CGCGCTTGGT	4320
TAACTTGGTC AAAGGAAGCG CGTGGCATGT AACATTCCCC ACGAACTGTG ATATCTAGTT	4380
CTTCTGGCAA AGTCAAAGG ATGTCCTTAA CACGCTTGAG GTTTTCTGTG ATATTTTCAC	4440
CAATTGAACC ATCTCCACGT GTTACCCAG CAACCAAAT CCCCTTTTCA TAAGTCAGCG	4500
AGATAGATAA GCCATCGATT TTCAGCTCAC AAATATAGGT CGGATGAGCC ACTTCTTAC	4560
GAACACGCGC ATCAAAAGCA TCTAGTCTCT CACATGAAAA AGCATCTCTG AACTATAAA	4620
GAGGATACTG ATGACTGTAT TTTTCAAAC CATCTAAAAC CTTGCCACCA ACACGATGAG	4680
TCGGAAGTGC TGCTAGCACT TGCTCTGGAT AAGCAGTTTC TAACTCGACC AACTCACGGT	4740
AAAGGCGGTC ATACTCACTG TCTGAAACCG AGGGATTATC GCTGGTATAG TACTCAGTCG	4800
CATAGCGATT GAGCAAAGCG ACTAACTCAT TCATTCTTTT ATTCATAAGA CCATTTTACC	4860
ATAAAACAAG CCCTCCTCAC AAACGAGAAG GCGGAAAAA ACACCTAGTT TGAAATTATT	4920
TTTGAAACTC AAGCAACCTT ATATCAATTT TTCAAATGA GTTCGAACAT ATCCGAGAGC	4980
TAAGAAATAT AAGGCTACAA CTCCAAGTCC AATAATCAAG AAAGAATAAA GATGGACACT	5040
TGGCAAGACT GTCATAAATC CTTTGCAAT AGGCATAAAT AGAATAGCTA AGGTAAAAAT	5100
TGTACTCAGT ACTCTTCCAA GAAATTCGCT CTCAACCTTG GTTTGTAATT GAGTAAAAAA	5160
GTGAATATTA AAAATCGTCA TAAACAATTC ACAAATAAA TTTCCAGAAA AGGAAAGAAA	5220
AGTTGGAAGT GGTAATCCCA TCATAAAAAC TCCGACACCT GTCAAAGCCA GTAAATCAA	5280
AAGATTATAA ATATTAGCTT TAATTTTACT AGCTAGAAGA GCCCCAATGA TGGAACCAAT	5340
AGCCCCATA GTTAAATAC TTGCATAGGC TCCTTCTGAC CCGTAAAGCT GATTGAAAA	5400
GGGAAGTAGA AATTCAAAAG CTGCAAAAAA GAAATTAACG CTGGAAGCTA CCAGCAAAAG	5460
GAAGAAAATT TCTTGCTGAT GCCAGATATA GTGTAACCCA TCCTTGATAT CTACAAAAAT	5520
ATCTCTCCCA GTAAAAGCCT TTTTCTCTTG AACTTTTGCT TCCTCTTTTG GAAGGAAAGC	5580
CACTAGAACA AAAGCAATGA AAAAAGTCAG CGAGTCTAGC AGTAGCGTCA TATGGAGACT	5640
TGCAAACTGT AAAACAAGGA AGGAAAGAAC AGGAGAGCTA ACACCTACAA CCTGCAAAAC	5700
CAGCTCTAAG CGAGAATTAT AGATCACAAT CTCATCTTTC TCCACCACTT CAGTTATGAT	5760

904

AGCTTTATTG	GCTGTGCGAG	AAAAGGCAAA	AGCAATAGCC	TGCACAATGT	TAGCAACAAT	5820
CAAAGCGCCA	ATCATCCAGC	TATCATTCCT	TATGAAAGAA	ATAGCCAGAC	AAAGAATCCC	5880
ACAAACAAGA	TCTGCCGTCA	TTAAATCTT	ACGACGAGAA	AAACGGTCTG	AAATAACTCC	5940
GCCAAAGGGA	TTGACGAGAA	TAGATGTGAC	GAGCTCAGAA	ATCTGATACA	TTCTTAAAC	6000
TGTCTGTCCT	ATAGTCCCCA	TAGAAGCCAA	CCAGACACTA	TTTCCATAAT	CATAGAGCAT	6060
ATTTCCCAT	TTATTGATAG	CCCCACGGCT	AATCAACTGC	ACTGCATAGC	GATTTCATTT	6120
AAAGCTCCTC	TCAAAATTTG	AAACTATTGT	ATCAAAACCG	AAAGGAGCTT	TTTATTTTTT	6180
CCCTTATTG	GGAAAATTAA	CTTTTGACAA	ATTTTTCGTA	GTGTTCTCGA	TAATAGGCTA	6240
CTTGCTCTGG	AAGACCTAAC	ACATCAAAAA	TATGCATGGC	CTCTTGCATC	TGCTTACAGC	6300
CTTCTTTACA	CTGTCTTTT	TGATATAAGG	CAAAACCTTT	TAAATAATGG	AAAACATTAC	6360
GCTCATAAAG	CTTAATACCT	TTGTCAATAA	TCTTCTCTGT	ATAAGCCTCA	AAATAGTTGG	6420
CATTATAAAA	AGAAGAATGC	TCTAAACAAT	GCTGGTAACA	ATTGAGGGCC	AAAATCAACA	6480
CTAATCTCTT	ATGGCGACTA	ATCTCTTGGT	AAAATTCCTC	CCTCTCCATA	ACTTCTCTAC	6540
CAATCCGAGT	GACATAGTCT	ACATCGTAGA	AACTATAGAG	GTTACCGAAA	AGAATCAACT	6600
CATACATGGT	CCATTCTTCT	GTTTTGAAGA	GATAATCTGC	TACCTTACCC	AAATCATCCT	6660
GCTTCATATC	ATAACTCGCA	TCTCTTTGAC	AAATCAGACC	TTGTAGCAAA	ATCCAGTTCA	6720
GCTCAAAATA	AAGGGGAGTC	GTCGAACTCT	TAGACTTTTC	AAGTTGTTCT	CTTTGAAGCT	6780
TTTGAAAACC	TGCAATATCG	TTTGAATAGT	AAAGTGGGAT	AATCTGTGCC	ATCATAGACA	6840
CATGTTTCATG	ATTATGAAAA	TTCCTTGCCT	TATCCATGAA	ATTTTCGATT	GTTACATGAA	6900
TGTTATCCAA	AATCTCAAAG	AAACGGGAGA	CTGCCAGGTC	AGACTCCCCA	AGCTCAAAGC	6960
GAGATAACTG	AGAGGTAGAG	CAGGATTCGC	CTGCTGCTTC	CTTTAAAGAA	TAATTTCCAC	7020
TTGTTGAAAA	TTACAGAAAT	ACTTTTCCAA	GATGTTCCAT	CTTTACACCT	GCTCTGATAA	7080
TTCTTCCCAC	TCAAGCATAG	CTTCTTCCTG	ACGATGGCTG	ATTTTGTCCA	GCTCAGCCTG	7140
TAATTCATG	AGTTTGTCCG	CATCGTTTGT	TTCCAACATT	TGTTTCAGAAA	TGGCTTGGCT	7200
TTGACTTTCT	AGCTCTTCAA	TTTCAGCTTC	TAGACTTTTCG	ATTTGTCCGA	TGAGTTTGCG	7260
AACTTCTTTT	TGACTTTCTT	TCTGGGCCTG	ATAGTCAATTG	ACTGGACTTG	CTTCTTTTGC	7320
TTGATTGCTA	GTTGAAGCTT	CCTCAGTCTG	ACTCATTTCT	GCTGTTGCTT	TCTTCTCAAC	7380
ATAGTAGTCG	TAATCTCCAA	GGTAGAGAGT	TGAACCATTC	TCAGACAATT	CCAAAACATG	7440
AGTTGCCACA	CGATTGATAA	AGTAACGATC	ATGACTGACA	AACAGCAAGG	TTCCATCAAA	7500
GTCAATCAAG	GCATTTTCTA	GCACTTCCTT	ACTATCAATA	TCCAAGTGGT	TGGTCGGCTC	7560



905

ATCCAGAATC AAAAAGTTAT TGTMTTCCAT AGACAATTTA GCTAAAAGCA AACGAGCTTT	7620
TTCGCCACCA GATAGCATGC CGACTGATTT TTAAACATCA TCTCCTGAGA AAAGGAAGGC	7680
TCCAAGACGG TTGCGGATTT CAACTTCTGG TGTCAGTTTG AAATCATTCC AGAGTTCATC	7740
CAGCACCGTA TTACTTGGTG TCAGCTTGCT TTGGGTTTGG TCATAGTAAC CAACCTCAAC	7800
ATTAGCGCCA AAGCGCTTTT CTCCTTGAT AAAAGGAATC TGGTCCACAA TAGACTTGAT	7860
AAAGGTTGAC TTGCCGATAC CATTTGGACC AACGATAGCG ACAGCATTCA TCTTACGAAG	7920
ATCTAGGTTA ATCGGTTGTG ACAAGACTTC CCCGTCATAG CCAACAGCTG CATTTTCAAC	7980
AGTCAAAACA ACATTGCCCG ACGTTTTTTC AGACTGGAAG GTCATGTTGG CTGATTTCTT	8040
GCCAGCTTCA GGCTTGTCGA AACGTTCCAT TTTTCCAGT TGTMTACGGC GAGATTGAGC	8100
ACGTTTAGTC GTTGAAGCAC GAACTAGATT GCGATTGACA AAGTCTTCCA GAGCAGCGAT	8160
TTCTTCTGT TGCTTTTCAT AGTTTTTTC CTCAGTAACT AGCTTTTGCT CCTTCAATTC	8220
GACAAAACGA GAGTAATTCC CCACATAGCG ATCCAAGGAA TGCTTGGTCA AATCTAGCGT	8280
AATTGTCGCA ACCTTGTCGA AGAAATAACG GTCGTGGCTG ACGATAATGA GGGCACCCT	8340
ATAGTTTACC AAGTAATTCT CTAGCCAGGC GATGGTTTCA ATATCCAAGT GGTTAGTTGG	8400
CTCGTCCAAG ACCAAGAGAT TGGGCTTTTC AAGGAGCATT TTGGCAAGTG CCAAACGAGT	8460
ATTTTGACCA CCAGAAAGCT CAGCAATTTT CATCTGCCAC ATAGACTCGT CAAACTTGAA	8520
TCCATTCAAA ATCGCTCGAA TATCAGCTTC ATAGGTAAAG CCACCTGCTT GGCGAAAATT	8580
CTCAGATAAG CGGTCAATAT CTGACATCAG TTTATCCAAA TCCTCACCAG ACTTTTCACC	8640
CATCTCCAGC TCCATCTGAC GCAGTTGTCT CTCGTCCTGA CGCAAATCAT TAAAGACATC	8700
AAGCATTTC ACGTAGATGG TATTTTCAGA CTCAAACGG CTATCTTGGG CTAGGTAAGA	8760
CAGAGAAATA TCTTTTCTT TATTGATTTT TCCGCTAGTT GGCTCCTCTT CTCCAATAA	8820
AATCTTCAAA AGAGTAGACT TACCTGCACC ATTTTCCCA ACAAGAGCAA TCCGATCTCG	8880
TTTATCAACC TGCAGGTTGA TATTATCGAA AAGAACCTCT CCTGCAAAAG AACGTTCAAT	8940
TTTATTAGCT TGTAATAATA TCATACAAGT AGTATAGCAT GTTCCCTAA GGCATTCAAG	9000
ATAATCGTAA GTCTTTTAGT ACAACTTTTA TAACATAAAA TAACTAAAT TATGTATATT	9060
TTATATTAGA TTACTTCACT ATCTTGTTGG ATTTTCTAAC CAGCTAATCT TGTTTCAAAT	9120
AGTTATCGCA CAAGTCTATT ATTTAATTCT TTTTCATCATT TACGTACGTA TAGCAGATTG	9180
AAATAAGATG AGAACAAATC GATTGGGAAA GTAAAATTAA TTTCTATAAA TGTTTTAGCA	9240
ATTGTTTCGT ACTATTTTAG ATTCAGTCTA CTATATACAA TATTTTCGGA ACATTCAACT	9300

906

TTTAACTCT ATTTATTACT AGATTTTCATA ATTAAAAAAC CTA CTGACCA AGCTAGAAAG	9360
CTTGATACAA TAGGCTTTTT AAAGACTGAT TATTTAACAG CGTCTTTAAG AGCTTTACCA	9420
GCTTTGAATG CTGGTACTTT AGAAGCTGCA ATTGTCATTT CTTTACCAGT TTGTGGGTTG	9480
CGACCTTTAC GTTCTGCGCG CTCACGAACT TCAAAGTTAC CAAAACCGAT CAATTGAACT	9540
T	9541

## (2) INFORMATION FOR SEQ ID NO: 133:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3502 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

TTGACTATCC TATCATGCTT TCTAAGGTCT ACTCAAGAAA ATCATTTTCA AGTTTTCACA	60
CCTTTCTCAA AAAAGTTAAA AAATTTTCTC AAAAACGCTT GACTCTGACC TAAGGCGAAG	120
GGTTATACTA TCATTGTAAG GAGGAAATCA TGTACCATAT AAAAGAAGCT GCGCAGCTTT	180
CAGGTGTCTC TGTCAAGACC CTGCATCACT ATGACAAGAT AGGACTCTTG GTCCCCTTAA	240
AGTCGGA AAA CGGCTATCGA ACCTACAGTC AAGAGGATTT GGAACGCCTT CAGGTCATTC	300
TTTACTACAA ATATCTAGGC TTTTCTTTAG AGAAAATAGC AGAGCTGTTA AAGGAAGAAA	360
GGACAGATT ATTGCCCCAT TTGACTAGGC AGTTGGACTA TCTAACTCGC GAAAGGCAAC	420
ATCTGGATAC CTTGATTTCC ACCTTGCAAA AACTATTCA AGAACAAAA GGAGAAAGAA	480
AAATGACCAT TGAGGAAAAA TTCACGGGAT TTAGCTATCA AGACAATCAA AAATACCACC	540
AAGAAGCGGT AGAGAAATAT GGTCAGAAG TCATGGGACA AGCGCTCGAA CGCCAAAAAG	600
GTCACGAAGA CGAGGCTACG GCCGCCTTTA ACCAAGTCTT TCAAACCTTG GCACAAAATC	660
TTCAAGTTGG TTTACCTGCA ACAGCAACCG AAAACCAGGA GCAAGCAGCC AAGCTCTTGC	720
AAGCCATTCG CACTTATGGA TTTGACTGCT CTATTGAGGT ATTCGGTCAT ATCGGTAAAG	780
GTTACGTCTA CAACCCAGAG TTTAAGGAAA ACATTGACAA GTTTGGTTCT GAAACAGCCC	840
AGTACACGTC AGATGCCATT GCGGTTTACG TTCAGACAAA TGCAGAATAA ATAGGCTAGG	900
AATTTCTAG CCTATTTTTT ACTTCAAATC ATAAAGCCAG TCGTCACCGT TTTTGTAGTA	960
AAAGAATTCA CTGAGATCTT CTTCTAGAAA CACACGAAGC ATATCAGACA TATCATCGGT	1020
TGCAAGTTTT AGATGAGAAA GATTTTCAAA GTCCTCCAC CAACTTTCC CTTCTGTGA	1080
AGACTGGAGT TCACCAGTAA AGTGTTCTGT CTTGTAAAAA AGGACGACAT AACGATAATC	1140

907

CTTGTCGTCA TACCAGTTTT TGATACCACA GAGTTGGGGT TTGGAAATGA TCAGACCAGT	1200
TTCTTCTTTC ACTTCACGAA TGACAGCATC GACAAAGGAT TCGCCACGTT CAACATGACC	1260
ACCAGGAAAA GTAATGCCAG ACCAGTCGGG ATTAACCTCG TCTTGGACCA GGACCTTATC	1320
TCCGTTTTTA ATCATACACA TGTTAACAAA TTCCACTGCC TCTCTTCTGT TCATTCTTCA	1380
CAACCTTTAA TCTTTAATCA TAATGCAGAC TTCCCGCCAC CCAGCCGGTA CAGAGGGCAG	1440
AAGTGATGTT AAAGCCACCC GTGTGGGCAT TGATATCCAT AACTTCGCCT GCAAAGTGGA	1500
GGCCAGGTAC CAGCTTACTT TCAAGGGTTT TAGGATTGAT TTCCTTGAGA CTGACTCCAC	1560
CCTTGGTAAAC AAAGGACTTT GCAAGGGACA TTTTCCAGT TACAGGAATT TTAAGTTCTT	1620
TAATGGACTG GACAAGTTGT TCTCGTTCCT TTTCAGTCAG TTGTTTGACT TTTTCAGGAT	1680
ATCCTTGATC AAAAAATTCG GCCAAGCGTT CTGGTAACAA GGTTTTTTAA GCGTTTTTCA	1740
AGGATTTTTT CCGATTTTCT TCTAGAAATG TAACCAAGTC CTTCTCAGAA AGTTGAGGCA	1800
AAACATCGAG TGAGAGAACC TCCCCACCTT TGACAAAGCT AGACATGCGT AGGGCAGCAG	1860
GACCTGACAA ACCAAAGTGG GTAAAGAGTA AATCATGAGT GATGACATGC TTACCATAAC	1920
TTAGGGTCAC ATCGTCCAGA GAAATACCTT GTAAGGCTTT ATGTGGAAAA TCTGTTAATA	1980
AAGGACTTTC AGCAGCCTCA AGATCGGTGA TGGTATGCTT AAAATGGCGA GCAATCTCGT	2040
GACCAAAACC AGTCGAACCA GTCGAAGGAT AAGACTTACC ACCTGTTGTG ACAATGAGTT	2100
TCTCACAAGT GAAGGTTTGA TCCGCTGACT TAAGGACAAA CTGGTCATCT ACTTTTTTAA	2160
CAGAAACGAT TTCTATTGTA GTAGCAACTT GACCACCTAG TTCGGTGATT TTCTTTTCCA	2220
AAGCTTCGAT AATAGTCCGA GACTTGTCAC TGGCTGGAAA GACCGCTCCG TGGTCTTCGA	2280
CCTTAAGTTT AACACCATTT TCTGTAAAAA AGTTGATGAT GTCATGATTA TCGAACTGGG	2340
AGAAAACACT GTAAAGAAAG CGTCCGTTTC CAGGAATTCC AGCTAGCAGG TTGTCTAAGC	2400
TACCATTGTT GGTACATTG CAACGTCCCC CACCAGTCCC AGCTAATTTT TTTCCAAGTT	2460
TCCGATTTTT TTCGATGAGG AGGGTTTTCT GTCCATAAAA GCTACTGGAA ATCGTAGCCA	2520
TCATACCAGC AGGTCCCCCA CCGATGACAA TAGTATCAAA ATGTTTCATA GCTCTATTGT	2580
ACCACAAAAA AACAAGAGAT GATGGTCACC TCTTGTCAGG AATGCAATTA ATCAATTTCA	2640
TAGCCCATCA GCAAACCGCC CTCTTCTGCA TAGAACTGC AGAGACCAGA GGTGGGTAGA	2700
ATTTTAATAT CCGCTTGTTG GAAGGTTTCA CGGATTCGCT CTGAGAGCTG TTGACAACAT	2760
TTTTCGTTAT TCGGTTGGGC CATGACAATA CGGCCACCAG CATATCCAGC TTTTACTAAC	2820
TCATCATAGG CAGCTTGAAC TGATTTCTTT GATCCCTTGT CTTTTTGTAG CAATTCGAGA	2880

908

GTCCCAGTTT CACTAGCTTT TCCGACCATA CGAATGTTGA GAAGGCCAAC GACCGTACCG	2940
ATAAGCTTGC TCAAACGGCC GTTCTTCACC AAGTTATCGA CTTTGGCTAG GACAAAGAGC	3000
AACTTAGTTT TTTCTTGATA GCGGTGATA GCTTCAACCA CTTCTTCAA AGACAAGCCC	3060
TGGTCAATCA AGTCATTCAA TTTTCTACG AGTAGGTCAA CTTACCACC AGCAGATAAA	3120
CTATCAATCA CATGAATCTT AGTGTACGGA TGGTCTTCCA GATAAATATT CTTGCTAGT	3180
TGAGCACTAT TGTGACTGCC AGAAAGGGTA CCTGTGATGG TTACTAGGAA AATGTTTTTG	3240
GCACCTTCAA ATGCTCGCAA ATAGTCATCT GGGCTTGGAC AAGCCGATTT TGAAGCTTCT	3300
GCAGTTGCAT ACATGGTTTC CATCATTTGG TCAATATCGA GACTGGCGTC ATCAACAAAG	3360
ACCTGATCAG CTAATTGAAT GGTAAAGGG ACACCTACAA AGGTGTGTG AATAGCTGGT	3420
GTTGGCAGTT GACGATAATC ACAACCAGAG TCAGCAATAA TCTTCCAAGT CATAGAAATT	3480
CTCCATCTTT GTCAGGAACG AT	3502

## (2) INFORMATION FOR SEQ ID NO: 134:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12665 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

CGATTGATTT TTTTAAAGCG TTCGATAGAG AATGAGAAAC GAATCCTTAG CAATGGCGGG	60
AAAGAATTG GAGTTGAGAA TACAAAACGA TTAATATGG CTCATATTGT TTTTATCTC	120
TCTTGCTTGG TTGAGGCAAT GGTGCACAAG ACAATTTTGG ATGGCATGGG CATGGTTGGT	180
TTAGTCTTGC TTATTTTTTC TATGCTGATG TTGATGTTGG TGATTCATT GTTGGGAGAT	240
ATTTGGACAG TGAAGCTTAT GCTTGTCAAT AATCACAAAT ATGTAGATCA TATCTTGTTT	300
AGGACAGTAA AACACCCTAA TTAATTTTTA AATATTCTTC CTGAGTTGAT TGGCTTGACC	360
TTGTTGAGTC ATGCTTATGT GACTTTTGTT TTAGTTTTTC CAGTTTATGC AGTTATTTTG	420
TATCGACGAA TAGCTGAAGA GGAAAAGCTA TTACATGAAG TTATAATCCC AAATGGAAGC	480
ATAAAGAGAT AAATACAAAA TTCGATTAT ATACAGTTCA TATTGAAGTG ATATAGTAAG	540
GTTAAAGAAA AAATATAGAA GGAAATAAAC ATGTTTGCAT CAAAAAGCGA AAGAAAAGTA	600
CATTATTCAA TTCGTAAATT TAGTGTGGA GTAGCTAGTG TAGTGTGTC CAGTCTTGT	660
ATGGGAAGTG TGGTTCATGC GACAGAGAAC GAGGGAGCTA CCCAAGTACC CACTTCTTCT	720
AATAGGGCAA ATGAAAGTCA GGCAGAACA GGAGAACAAC CTAATAAACT CGATTCAGAA	780

909

CGAGATAAGG CAAGGAAAGA GGTGAGGAA TATGTAAAAA AAATAGTGGG TGAGAGCTAT	840
GCAAAATCAA CTAAAAAGCG ACATACAATT ACTGTAGCTC TAGTTAACGA GTTGAACAAC	900
ATTAAGAAGC AGTATTTGAA TAAAATAGTT GAATCAACCT CAGAAAGCCA ACTACAGATA	960
CTGATGATGG AGAGTCGATC AAAAGTAGAT GAAGCTGTGT CTAAGTTTGA AAAGGACTCA	1020
TCTTCTTCGT CAAGTTCAGA CTCTTCCACT AAACCGGAAG CTTACAGATAC AGCGAAGCCA	1080
<hr/>	
AACAAGCCGA CAGAACCAGG AGAAAAGGTA GCAGAAGCTA AGAAGAAGGT TGAAGAAGCT	1140
GAGAAAAAAG CCAAGGATCA AAAAGAAGAA GATCGTCGTA ACTACCCAAC CATTTACTTAC	1200
AAAACGCTTG AACTTGAAAT TGCTGAGTCC GATGTGGAAG TTAAAAAGC GGAGCTTGAA	1260
CTAGTAAAAG TGAAAGCTAA CGAACCTCGA GACGAGCAAA AAATTAAGCA AGCAGAAGCG	1320
GAAGTTGAGA GTAAACAAGC TGAGGCTACA AGGTTAAAAA AAATCAAGAC AGATCGTGAA	1380
GAAGCAGAAG AAGAAGCTAA ACGAAGAGCA GATGCTAAAG AGCAAGGTAA ACCAAAGGGG	1440
CGGGCAAAAC GAGGAGTTCC TGGAGAGCTA GCAACACCTG ATAAAAAGA AAATGATGCG	1500
AAGTCTTCAG ATTCTAGCGT AGGTGAAGAA ACTCTTCCAA GCCCATCCCT GAAACCAGAA	1560
AAAAAGGTAG CAGAAGCTGA GAAGAAGGTT GAAGAAGCTA AGAAAAAGC CGAGGATCAA	1620
AAAGAAGAAG ATCGCCGTAA CTACCCAACC AATACTTACA AAACGCTTGA ACTTGAAATT	1680
GCTGAGTCCG ATGTGGAAGT TAAAAAGCG GAGCTTGAAC TAGTAAAAGA GGAAGCTAAG	1740
GAACCTCGAA ACGAGGAAAA AGTTAAGCAA GCAAAAGCGG AAGTTGAGAG TAAAAAGCT	1800
GAGGCTACAA GGTTAGAAAA AATCAAGACA GATCGTAAAA AAGCAGAAGA AGAAGCTAAA	1860
CGAAAAGCAG CAGAAGAAGA TAAAGTTAAA GAAAAACCAG CTGAACAACC ACAACCAGCG	1920
CCGGCTCCAA AAGCAGAAAA ACCAGCTCCA GCTCCAAAAC CAGAGAATCC AGCTGAACAA	1980
CCTAAAGCAG AAAAACCAGC TGATCAACAA GCTGAAGAAG ACTATGCTCG TAGATCAGAA	2040
GAAGAATATA ATCGCTTGAC TCAACAGCAA CCGCCAAAAA CTGAAAAACC AGCACAACCA	2100
TCTACTCCAA AACAGGCTG GAAACAAGAA AACGGTATGT GGTACTTCTA CAATACTGAT	2160
GGTTCAATGG CGACAGGATG GCTCCAAAAC AATGGCTCAT GGTACTACCT CAACAGCAAT	2220
GGCGCTATGG CGACAGGATG GCTCCAAAAC AATGGTTCAT GGTACTATCT AAACGCTAAT	2280
GGTTCAATGG CAACAGGATG GCTCCAAAAC AATGGTTCAT GGTACTACCT AAACGCTAAT	2340
GGTTCAATGG CGACAGGATG GCTCCAATAC AATGGCTCAT GGTACTACCT AAACGCTAAT	2400
GGTTCAATGG CGACAGGATG GCTCCAATAC AATGGCTCAT GGTACTACCT AAACGCTAAT	2460
GGTGATATGG CGACAGGTTG GGTGAAAGAT GGAGATACCT GGTACTATCT TGAAGCATCA	2520

910

GCTGCTATGA AAGCAAGCCA ATGGTTCAAA GTATCAGATA AATGGTACTA TGTCAATGGC	2580
TCAGGTGCCC TTGCAGTCAA CACAAGTGA GATGGCTATG GAGTCAATGC CAATGGTGAA	2640
TGGGTAACT AAACCTAATA TAACTAGTTA ATACTGACTT CCTGTAAAGAA CTCTTTAAAG	2700
TATTCCCTAC AAATACCATA TCCTTTCAGT AGATAATATA CCCTTGTAGG AAGTTTAGAT	2760
TAAAAAATAA CTCTGTAATC TCTAGCCGGA TTTATAGCGC TAGAGACTAC GGAGTTTTTT	2820
TGATGAGGAA AGAATGGCGG CATTCAAGAG GCTCTTTAAG AGAGTTACGG GTTTTAAACT	2880
ATTAAGCCTT CTCCAATTGC AAGAGGGTTT CAATCTCTGC CAGGGTGCTG GCTTGCGAAA	2940
TGGCTCCACG GAGTTTGGCA GCGCCAGATG TTCCACGGAG ATAGTGAGGA GCGAGACCGC	3000
GGAATTCACG AACTGCGACG TTTTCTCCTT TGAGGTTAAT CAATCGTTTC AAGTGTTCTG	3060
AGGCGATCTT CATCTTGTCT TCAAAGGTCA AATCAGGTAG GATTTCTCCT GTTTCAAAGT	3120
AATGGTTGAT TTGGTTGAAG AGGTAAGGAT TTCCCATGGC AGCTCGGCCA ATCATGACTG	3180
CGTCAGCACC AACTTCTTCG ATGCGTTGCT TGGCTTCTTG GACAGTACGG ATATCACCGT	3240
TGGCGATGAA TGGAACTTTG GTTAGAGCTT GGGCAACCTT GTAAAGGGTC TCAAGGTCTG	3300
CGTGGCCAGT ATACATTTGT TCACGGGTAC GGCCATGCAT GGCGAGGGCA GAAACACCTG	3360
CAGCTTCAGC AGCGAGAGCA TTTTCTACTG CAAGAGATGG GTCCGCCCAG CCGGTACGCA	3420
TTTTGACAGT AAGTGGGATA TCAAGGACAG ACTGGACCTT GTTGATGATG GAGTAAATCT	3480
TGTCTGGATC CTTGAGCCAC ATAGCACCAG CTTCTTCTT CACGATTTTG TTGACAGGGC	3540
AGCCCATGTT GATATCGACG ATATCGGTCT TGGTGTMTTC TTGGATGAAT TCTGCTGCGC	3600
GTGCTAGGCT GTCTTCATCG CTACCAAAAA GTTGGATAGA GACAGGGTTT TCGCCCTCAT	3660
CGATATGAAG CATGTGCAGG GTTTTTTCGT TGTGTATTG GATTTCCCTG TCAGAGACCA	3720
TTTCCATTAC AACGAGTCCA GCTCCGAGCT CCTTTGCCAT AGTACGAAAG GCTGAGTTGG	3780
TCACGCCAGC CATAGGCGCT AAAACGGTAC GATTGGGAAT CTCAATATTG CCAATCATAA	3840
AAGGTGTATT AAGATTTGTC ACGAATGAGT TCCTCCAGGT CCTTTTCATC AAAGTTGTAA	3900
GTAGTTTGGC AGAATTGACA AGTGATTCTT GCGCCGTGGT CTTCTCTTTT CATTTCTCTG	3960
AAGTCTGAGC TTGGAAGGCT GGCAAGAGCG TTCATAAAGC GTTCATGGCT ACAGTCACAT	4020
TGGAAACGGA TTTCTTCTTC AGAAAGACGC TTGTAGGCTT CGTCCCCGTA GATAGCCTTG	4080
AGGAGGGCTT CGATATGGTC GTCGCTTTCG AGAAGACTAG AGATAGCTGG CATTTCTTGG	4140
ATGCGTTTTT CAAAGCGAGC AATCTCTTCT TTCTTGGCTC CTGGCAAGAC TTGAACTAGG	4200
AAACCACCTG CAACCTTGAC CTTGTCTTCC TCGTCCAAAA GGACATTGAG GCCGACCGCT	4260
GAAGGCGTTT GTTGGCTTTC AGTAAGGTAA AAGGCAAGGT CTTACCCGAT TTCTCCAGAG	4320

911

ATGAGGGGAG TTATAGAGTT GTAAGGATTT CCAGTACCGT AGTCTGTGAT AACGAGGAAT	4380
TGACCATTTC CAACAAAAGG TCCGACTAGG ACTTCACCAG TCGCAGTCTT TTTGATGTCA	4440
ACACCAGGAT TTTGAACATA GCCTTTGACG TTCCCCTTGG TATCAGCGAC GGTGATAATA	4500
GCACCTAGAG AGCTAGATCC CAACACCTTA ACTGTAAGTT TGGTATTTCC TTTTTCATTG	4560
GCTGCGAGAA TCTGGCTAGC GATAAGAGTT CGACCAAGCG CTACAGTTGA GCTAGCTTGG	4620
GTTTGATGTT TTTCTTGAGC AGTGCGGACG GTTTCAGTGC TATCAAGGAC AAAAGCACGA	4680
AAGGCTCCGC TTTCTGATAT AGTTTAAATA ATTTTATCCA TAGCTACTAT TTTAGCATAA	4740
AAATGCCCAA AGGGGGAGCC GTGTGTTTAC TGATTTTCAG GATAATGGAC CAGGAAATCA	4800
GCATGAAAAT AAAAAGAGAA ACAGATTATT TTAGCATTTG TCAGATTTAT GCTATGCTTA	4860
AGGTAGAAAA TGAAAGGGAT AACAAATGTA TTTAGGAGAT TTGATGGAGA AAGCCGAGTG	4920
TGGTCAATTT TCAATACTTT CCTTCTATT ACAAGAGTCT CAGACGACCG TCAAGGCTGT	4980
AATGGAAGAA ACAGGATTTT CAAAAGCAAC CCTAACCAA TATGTCACCC TGCTCAATGA	5040
CAAGGCTTTG GATAGTGGCT TAGAGCTGGC TATTCACCTA GAAGATGAAA ATCTGCGTCT	5100
GTCTATCGGT GCAGCTACCA AGGGGAGAGA TATTCGGAGC TTGTTTTTGG AGAGTGCTGT	5160
TAAATACCAG ATTTTGTTTT ATCTTCTCTA CCACCAACAG TTTTtagccc ATCAGCTGGC	5220
TCAAGAATTG GTGATTAGCG AGGCTACGCT TGGTCGTCAC TTGGCTGGTT TAAATCAGAT	5280
TTTGTCAGAA TTTGATTTAT CCATCCAAAA TGGCCGTTGG CGAGGTCCAG AGCATCAGAT	5340
TCACTATTTC TATTTCTGTC TTTTCCGAAA GGTCTGGTCG AGTCAGGAAT GGGAAAGTCA	5400
CATGCAGAAA CCAGAGAGAA AACAGGAGAT TGCCAATTTA GAGGAAATCT GCGGTGCAAG	5460
TTTGTCTCGG GGGCAGAAAT TGGACTTGGT TCTCTGGGCT CACATCAGTC AACAACTCT	5520
TCGGGTCAAT GCTTGTCACT TTCAAGTCAT AGAAGAGAAA ATGCGAGGGT ATTTTGACAA	5580
TATCTTTTAT CTTGTTTTGC TGAGAAAGGT TCCGTCCTTT TTTGCTGGGC AACATATTCC	5640
ACTAGGAGTT GAGGATGGTG AGATGATGAT ATTCTTCTCT TTTCTCCTAT CTCATCGCAT	5700
TCTTCCTCTT CATACTATGG AGTATATTCT TGGTTTTGGA GGGCAGTTGG CAGATTTACT	5760
GACGCAATTG ATTCAAGAAA TGAAGAAGGA GGAACATTG GGGGATTATA CAGAGGACCA	5820
TGTCACCTAT GAACTCAGTC AGCTTTGTGC TCAAGTCTAT CTCTATAAGG GCTATATTTT	5880
ACAGGATCGC TACAAGTACC AGTTAGAGAA TCGTCATCCA TATTTACTGA TGGAACATGA	5940
TTTTAAAGAG ACAGCAGAGG AGATTTTCA TGCTCTACCT GCTTTTCAAC AGGGGACAGA	6000
TTTAGATAAG AAGATTCTCT GGAATGGCT CCAGTTAATC GAATATATGG CTGAAAACGG	6060

912

TGGCCAGCAT ATGCGGATTG GTCTGGATTT GACATCTGGT TTTCTTGTCT TTTCAAGGAT	6120
GGCAGCCATT TTGAAACGGT ATTTGGAATA CAATCGTTTT ATTACCATTG AAGCTTATGA	6180
CCCTAGTCGG CATTATGATT TGCTGGTTAC CAATAACCCG ATTCATAAGA AGGAACAGAC	6240
ACCAGTCTAT TATTTAAAAA ATGACTTGGA TATGGAGGAT TTGGTAGCGA TTCGCCAGTT	6300
ATTATTCACT TAAAAGGCTT GGTTAATCCA GGTCTTTTTT GTGAAATTCA CACAATCTCC	6360
TCACATTTTT TTA AAAAATTA AAAAAAGTTG ATAAACAAGA AAGCGCTTTA TTTGTATAC	6420
TAGTAAGTGT AAAGAGGAAA CACCTCAAGA TCTTTATCAG GAGGACAGTA CATGTCACAA	6480
GAAAAATACA TCATGGCCAT TGACCAGGGA ACTACAAGTT CTCGTGCCAT CATTTTCAAC	6540
AAAAAAGGGG AAAAGGTTAG CTCGACTCAA AAAGAGTTTA CCCAGATTTT CCCTCAGGCA	6600
GGTTGGGTTG AGCACAATGC CAATGAAATT TGGAACTCTG TTCAGTCAGT TATTGCGGGT	6660
GCTTTCATCG AAAGTGGTGT CAAGCCAAAT CAAATCGAGG CAATCGGGAT TACCAACCAA	6720
CGTGAAACAA CGGTGTCTG GGATAAGAAA ACAGGACTTC CTATCTACAA TGCTATCGTT	6780
TGGCAGTCAC GCCAGACAGC ACCTTTGGCT GAGCAACTAA AAAGCCAAGG TTATGTGGAA	6840
AAATTCATG AAAAGACTGG TTTGATTATT GATGCTTACT TCTCTGCTAC CAAGGTTCGT	6900
TGGATTTTGG ATCATGTAGA AGGTGCTCAA GAGCGAGCAG AAAAAGGGGA ATTGCTCTTT	6960
GGTACTATCG ATACTGGTT GGTTTGGAAT TTGACTGACG GTGCGGCTCA CGTGACTGAC	7020
TACTCAAATG CAGCTCGTAC CATGCTTTAT AACATTAAAG AACTCAAATG GGATGATGAG	7080
ATTTTGGAAT TCCTTAACAT TCCGAAGGCT ATACTTCCAG AAGTTCGTTT TAACTCCGAA	7140
ATCTACGGCA AGACAGCTCC ATTCCATTTC TACGGTGGAG AGGTGCCAAT CTCAGGTATG	7200
GCTGGGGACC AACAAGCAGC CCTCTTTGGA CAGTTGGCTT TTGAGCCAGG TATGGTTAAG	7260
AATACTTATG GAACAGGCTC TTTCATCATC ATGAATACTG GGAAGAGAT GCAGTTGTCT	7320
GAAAAACAAC TCTTGACAAC CATTGGTTAC GGAATCAACG GTAAGGTTTA TTATGCCTTG	7380
GAAGGTTCTA TCTTCATCGC AGGAAGTGCT ATTCAGTGGC TTCGTGACGG TCTTCGCATG	7440
GTTGAAAATT CACCAGAATC TGAAAAATAC GCTCGTGATT CTCACAACAA CGATGAAGTT	7500
TATGTCGTTT CAGCCTTTAC AGGTCTAGGC GCTCCATACT GGAACCAAAA TGCTCGTGGT	7560
TCCGCTCTTG GTTTGACTCG TGGAACAAGC AAAGAAGACT TTATCAAGGC GACTTTGCAA	7620
TCTATTGCTT ATCAAGTGCG TGATATCATC GACACCATGC AAGTGGATAC TCAGACCGCC	7680
ATTCAAGTAC TGAAGGTGGA TGGTGGTGCA GCCATGAACA ACTTCCTCAT GCAGTTCCAG	7740
GCGGATATTT TAGGCATTGA CATTGCACGT GCTAAAAACC TGGAAACAAC AGCTCTAGGA	7800
GCGGCCTTCC TAGCAGGTTT GTCAGTAGGG TACTGGAAAG ACTTGGACGA GTTGAAACTC	7860



913

TTGAACGAGA CAGGAGAACT CTTTGAGCCA TCTATGAACG AATCTCGCAA GGAACAACTC	7920
TACAAGGGCT GGAAGAAGGC TGTGAAAGCA ACTCAAGTCT TTGCGGAAGT AGACGACTAA	7980
TACTGGCAGA ATAAAGCGAT TTATTTAGAA AGTGTGTAAA TATGGAATTT TCAAAGAAAA	8040
CACGTGAATT GTCAATPTAAA AAAATGCAGG AACGTACCCT GGACCTCTTG ATTATCGGTG	8100
GAGGAATCAC AGGAGCTGGT GTAGCCTTGC AGGCGGCAGC TAGCGGTCTT GAGACTGGTT	8160
TGATTGAAAT GCAAGACTTT GCAGAAGGAA CATCTAGTCG TTCAACAAAA TTGGTTCACG	8220
GAGGACTTCG TTACCTCAAA CAATTTGACG TAGAAGTGGT CTCAGATACG GTTTCTGAAC	8280
GTGCAGTGGT TCAACAAATC GCTCCACACA TTCCAAAATC AGATCCAATG CTCTTACCAG	8340
TTTACGATGA AGATGGAGCA ACCTTTAGCC TCTTCCGTCT TAAAGTAGCC ATGGACTTGT	8400
ACGACCTCTT GGCAGGTGTT AGCAACACAC CAGCTGCGAA CAAGGTTTTG AGCAAGGATC	8460
AAGTCTTGGA ACGCCAGCCA AACTTGAAGA AGGAAGGCTT GGTAGGAGGT GGAGTGTATC	8520
TTGACTTCCG TAACAACGAT GCGCCTCTCG TGATTGAAAA CATCAAACGT GCCAACCAAG	8580
ACGGTGCCCT CATTGCCAAC CACGTGAAGG CAGAAGGCTT CCTCTTTGAC GAAAGTGGCA	8640
AGATTACAGG TGTGTAGCT CGTGATCTCT TGACAGACCA AGTGTTTGAA ATCAAGGCC	8700
GTCTGGTTAT TAATACAACA GGTCCCTTGA GTGATAAAGT ACGTAATTTG TCTAATAAGG	8760
GAACGCAATT CTCACAAATG CGCCCAACTA AGGGAGTTCA CTTGGTAGTA GATTCAAGCA	8820
AAATCAAGGT TTCACAGCCA GTTTACTTCG ACACAGGTTT GGGTGACGGT CGTATGGTCT	8880
TTGTTCTCCC ACGTGAAAAC AAGACTTACT TTGGTACAAC TGATACAGAC TACACAGGTG	8940
ATTTGGAGCA TCCAAAAGTA ACTCAAGAAG ATGTAGATTA TCTACTTGGC ATTGTCAACA	9000
ACCGCTTCCC AGAATCCAAC ATCACCATTG ATGATATCGA AAGCAGCTGG GCAGGTCTTC	9060
GTCCATTGAT TGCAGGGAAC AGTGCCTCTG ACTATAATGG TGGAAATAAC GGTACCATCA	9120
GTGATGAAAG CTTTGACAAC TTGATTGCCA CTGTTGAATC TTATCTCTCC AAAGAAAAAA	9180
CACGTGAAGA TGTTGAGTCT GCTGTCAGCA AGCTTGAAAG TAGCACATCT GAGAAACATT	9240
TGGATCCATC TGCAGTTTCT CGTGGGTCTA GCTTGGACCG TGATGACAAT GGTCTCTTGA	9300
CTCTTGCTGG TGGTAAAATC ACAGACTACC GTAAGATGGC TGAAGGACCT ATGGAGCGCG	9360
TGGTTGACAT CCTCAAAGCA GAATTTGACC GTAGCTTTAA ATTGATCAAT TCTAAAACCT	9420
ACCTGTCTTC AGGTGGAGAA TTGAACCCAG CAAATGTGGA TTCAGAAATC GAAGCCTTTG	9480
CGCAACTTGG AGTATCACGT GGTTTGGATA GCAAGGAAGC TCACTATCTG GCAAATCTTT	9540
ACGGTTCAAA TGCACCGAAA GTCTTTGCAC TTGCTCACAG CTTGGAACAA GCGCCAGGAC	9600

914

TCAGCTTGGC	AGATACTTTG	TCCCTTCACT	ATGCAATGCC	CAATGAGTTG	ACTCTTAGCC	9660
CAGTTGACTT	CCTTCTTCGT	CGTACCAATC	ACATGCTCTT	TATGCGTGAT	AGCTTGGATA	9720
GTATCGTTGA	GCCAATTTTG	GATGAAATGG	GACGATTCTA	TGACTGGACA	GAAGAAGAAA	9780
AAGCAACTTA	CCGTGCTGAT	GTCGAAGCAG	CTCTCGCTAA	CAACGATTTA	GCAGAATTAA	9840
AAAATTAAGA	AAAAATAAAA	GAGGTGGAGG	GCAGCATTC	TTGTCGCCCG	TCCCTTCTTT	9900
TTAATGGAGA	CAGAAAGATG	ATGAATGAAT	TATTTGGAGA	ATTTCTAGGG	ACTTTAATCC	9960
TGATTCTTCT	AGGAAATGGT	GTTGTTGCAG	GTGTGGTTCT	TCCTAAAACC	AAGAGCAATA	10020
GCTCAGGTTG	GATTGTGATT	ACTATGGGTT	GGGGGATTGC	AGTTGCGGTT	GCAGTCTTTG	10080
TATCTGGCAA	GCTCAGTCCA	GCTTATTTAA	ACCCAGCTGT	GACCATCGGT	GTGGCCTTAA	10140
AAGGTGGTTT	GCCTTGGGCT	TCCGTTTTGC	CTTATATCTT	AGCCCAGTTC	GCAGGGGCCA	10200
TGCTGGGTCA	GATTTTGGTT	TGGTTGCAAT	TCAAACCTCA	CTATGAGGCA	GAAGAAAATG	10260
CAGGCAATAT	CCTGGCAACC	TTCAGTACTG	GACCAGCCAT	CAAGGATACT	GTATCAAAC	10320
TGATTAGCGA	AATCCTTGGA	ACTTTTGTTT	TGGTGTGAC	AATCTTTGCT	TTGGGTCTTT	10380
ACGACTTTCA	GGCAGGTATC	GGAACCTTTG	CAGTGGGAAC	TTTGATTGTC	GGTATCGGTC	10440
TATCACTAGG	TGGGACAACA	GGTTATGCCT	TGAACCCAGC	TCGTGACCTT	GGACCTCGTA	10500
TCATGCACAG	CATCTTGCCA	ATTCCAAACA	AGGGAGACGG	AGACTGGTCT	TACGCTTGGA	10560
TTCTGTGTGT	AGGCCCTGTT	ATCGGAGCAG	CCTTGGCAGT	GCTTGTATTC	TCACTTTTCT	10620
AGTTTATACT	CTTCGAAAAT	CAAATTCAAA	CCACGTCAGC	GTCCCTTAC	CGTACTCAAG	10680
TACAGCTTGC	GGCTAGCTTC	CTAGTTTGCT	CTTTGATTTT	CATTGAGTAT	TAGAAAACAA	10740
TTATGTTGAT	AGAGCTTGGG	CAAGAGCCCA	ATTTACAGCA	AAAATGAAGT	AAATCTTCTC	10800
ATAATAAAAC	GCATCATATC	AAGCACGAAA	ATTCACGAG	GTCAACTACA	GTCAGAAAGC	10860
TGAACAACAA	GCCAAAACGC	CCAAAAAAGG	CGGCAAAAAG	CAAGCACCTG	CAAGCAACGT	10920
GCCGAAATGG	TCAAATCCTG	ATTATGTCAA	CGAATTAGAC	CCAAAAATCG	TTGATATGCT	10980
AGTAGAATTT	CACAAGTCAC	AAGGCACTTT	GGAAACTCCC	GAGGCGCAAG	CAGAAATCGC	11040
CCAAAAACGT	GAAGAAATCG	AGCAAAGGAG	AGCTGAGCTT	GAGGGTAAAA	AACAAGAGCT	11100
TTTGAACCGC	TTGAACAAAT	AGAGTTTCGC	AAGTATTATG	CTTACAAATT	ACTTGAGCAA	11160
TTAACTAAAA	TATAAACCTT	GCCTTTATAT	CTAGGCAGGG	TTTATATTTT	AGAAATTCAC	11220
GTAGGTTGTT	ACGGTTTTTA	CATACCCAGT	ATAGTTTGAG	TTTCTATAGT	ATTCAGTGAT	11280
AAACTTCCAT	TTTCTTTGAG	CAACATGGAT	ATAAGTACTT	GTTATGTAGT	ATGGATATGG	11340
GCTTTGTGAA	TCCAAGTAAG	ACTGATAAGC	TTGTATACCA	AAATATGCTC	CACCAATTAT	11400

915

TGCACCCCAT GGACCCCCCA ATAAAGCACC TATCCTACCA ATCATATAAC TGATTCCAGC	11460
ACCAGTCATG AAGTTAGCGA ATGTGTTAGC TTGTTTATTC CCATGTATTG TGTGACGTA	11520
ATTCCAAACA TTAGGATCGT ATGATCTAAA AGATATATTT AGGTCGATTT CATTCCTTTG	11580
ATAAGCCATA TAAATGCCC CATTGATATA GACGCCGTCA GCACGTCGTT CAATAGTGTC	11640
TACACTTCCA TCTGGATTGA CAACCTCAAG AACTTCATCG CTTAAAAATAT TTACTTGCGT	11700
ATCTCCGAAC CGCACTGATG AGCCATTCTC AAAGTGAGCC TCACCAGATA CAACTTTAGA	11760
GTTTGCCGAT AAGCTATCAT CAGCAAAAAC AAACAAGCGA CGGGGAAATG CTAGACATAC	11820
AGAAAACAGA CATAACTAGC AAACACATGC ATTTAAACAT CTTAGACATA ACGGAAACTC	11880
CTTTGTATTT TTGATTTTTT TCAACTTTTA TTATACAATA AAACCAAATA AAAAGAAAGC	11940
GGTAACAATA TGCTTAATGC GAAAATTTTT TATATATTTT TATGTTTGAT CGTTATCGAA	12000
ACTACAGGCT TGTTGTTGTT GAAAAGAGGT CTCGAAATGG GTTATTTAGA CACAGAAGCT	12060
ATTATCCTCG CAGTTTTTTC ATTTGCTTTT TACAACCTAT GTTCATTCGC TTGGGTCTGC	12120
TCTACAATAA AAAACAATAA AAAATAAATA GACGTATTTT CAAAAAAAAC maAATGCATA	12180
TTTATATTAG CAAAACGACG ATTTAAATCG TCGTTTTTTT GTAGTACGAC GGGCATGTCG	12240
TATATCTGAG GTGTAAGTCC TCAGCCTGAC TATCGTGAGG TAGCAGGGAG AGGAAGGGAT	12300
AGCGAAATCG TGGCTCTACG AACAGGAACG TGATAGTAAG GCGTATATAG CGGATAAGGA	12360
GGCTTCAAAC TCTAAAGTCC AAAAAGGTAG TCGTAACCTA TATGTGTAAA TCACGAGAGT	12420
AATTGAATTC GGAATAAGGT TTGTGTGAAA AAGATAAATC TTTCTAGAGT CTAAAGACTC	12480
TGCGTCAGAT TTCCTATTTT CACTGTAAAC TTTTAACGTC CTCATATCTT GTATAAACGA	12540
GGAAAGATGT ACGACTTATC CCGTGAGGTT TCATGAGCGT GAAAGCGTAG TAACAACGAA	12600
TCATGAGAAG TCAGCCGAGC CCATAGTAGT GAGGAAACTT CCGTAATGGA AGTGGAGCGA	12660
AGGGG	12665

## (2) INFORMATION FOR SEQ ID NO: 135:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5305 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

CGCTAATCAC TACAATCATT TTATTGTACT TTTTCACTCT CAAGAAAAGC AAGAAGTATT

60

916

CATTTTAGTT TCATTTAGTA TTATTTTGCA TACCTAAAAT ACAGTAAAAA ATCAGTCATC	120
TTGGTATGCT CCTGCTTTCA CTATTCAACA CGTTTTTGAC TTATACTAGG CTCATTTCCTA	180
AAAGCATTAT ATAATAGTGA TATGAAACCA ACTAAACTAA ACAAGAAATA TAAGCAATAA	240
AAATTCGTTT AAAAGATCTT ACTAAAGCTA ATACTAAATA AAAATAAAAG AGTAAACTAG	300
GAAGTTTATT TCAAACAACC TAAAATACTG ATTTTCGGCT GAAGATAATA CTGGAGTGCA	360
AATTAATGGG GTTATAATAA ATAGCTGATA GCTTGTGTG GTTTTGGATT TTTTAAGAGT	420
AGATGAGTAT TAAACTATA AGGAGGACGA AGGTGGCTAA AAATTTAAAA TTAATAATTAG	480
CTCGGGTAGA GCGTGATTTA ACACAAGGTC AACTGGCAGA GGCTGTCTGGG GTCACACGCC	540
AGACTATTGG TTTAATAGAG GCGGGAAAAT ACAATCCCAG TCTCTCGCTC TGCCAGTCTA	600
TTTGCAGATG TTTAGGGAAA ACCCTAGACC AACTATTTTG GGAGGAAGAA GATGAAAAAT	660
AGATTTTATT ATTCTCAATT ACTAGACGAA AGAGAAGAAC AACTGTTCAA TAAAGCGGGC	720
TCTGAAAGTT TCTATATCTG CATTGCTTTG TCGCTCCTAT CTTATATCAT TTCAGTATTA	780
GCACCAAGCC TTTTAATTC TAATATGCTG CTAATCGTTA TCATCATAGG GACATTTTAC	840
TTTTTCAATC GTGCCCCTTA TCTGGGAGTG ACCTACTATG GTCGTTTTCA TTTTACGATT	900
TTGGGTTGTT TTTTCCTAAC CTTGGCTATT ACGGCTCTTT TGATGTTGCA GAATTATCAA	960
TTCAACATAG AAATTTATCA GCACAATCCT TTGAATTTTA AATACCTGTC TGCTTGGGTC	1020
ATTACTTATA TCATTTACCT TCCGTGGATC TTTATTGCCA ATCTTGGTCT TAAGAGCTAT	1080
GGCGAATGGG CTCAGAAAAA ATTTGAACAA GATATGGATG AATTGGAGAG TGGAGAATAG	1140
CTTGTTACTC TTTTCTCAAT CCAGCTAAAA TGTGATATAA TAGTACTAAT TTATTGGAAT	1200
ACATGAAAGT TCTTGAAAA TTTTCATGGGT TTCTAGCTAA GGAAGTAGGA AAAGTATGTA	1260
TCCAGATGAT AGTTTGACAT TGCACACGGA CTTGTACCAG ATCAACATGA TGCAGGTTTA	1320
CTTTGACCAA GGGATTACCA ATAAGAAGGC GGTCTTTGAG GTGTATTTCC GCCAACAGCC	1380
TTTTAAGAAC GGCTATGCGG TTTTGCAGG TTTAGAAAGA ATTGTAAGT ATCTTGAAGA	1440
CTTGCGTTTT TCAGATAGTG ATATAGCCTA TTTGGAGTCG CTTGGTTATC ATGGGGCGTT	1500
CTTGGATTAC CTTCCGAATT TCAAGTTGGA GTTGACCGTT CGTTCTGCCC AAGAAGGGGA	1560
TTTGGTTTTT GCTAATGAAC CGATTGTGCA GGTGGAAGGA CCTCTAGCCC AATGTCAGTT	1620
GGTCGAAACG GCTCTTTTGA ACATCGTCAA CTACCAGACT TTGGTGGCGA CGAAGGCAGC	1680
TCGTATTTCG TCGGTTATCG AAGATGAACC CTTGATGGAG TTTGGGACAC GTCGGGCTCA	1740
AGAAATGGAT GCGGCCATCT GGGGAACACG CGCAGCTGTG ATTGGTGGCG CCAATGGAAC	1800
CAGCAACGTG CGTGCGGGTA AGCTCTTTGA CATTCCTGTT TTGGGAACCC ATGCCCATGC	1860

917

CTTGGTACAG GTTTATGGCA ATGACTATGA AGCTTTCAAG GCTTACGCTG CGACCCACAA	1920
AAATTGTGTC TTTCTTGTGG ATACCTATGA CACCCTTCGC ATCGGTGTAC CAGCTGCCAT	1980
TCAGGTGGCG CGTGAGCTGG GTGATCAGAT TAACTTTATG GGTGTGCGGA TTGACTCTGG	2040
GGATATTGCC TACATTTCTA AGAAAGTCCG TCAGCAACTG GATGAGGCTG GATTACAGA	2100
GGCTAAGATT TATGCTTCTA ATGATCTAGA TGAAAATACC ATCCTTAACC TCAAGATGCA	2160
AAAGGCCAAG ATTGATGTCT GGGGTGTGGG TACCAAGCTG ATTACAGCCT ATGACCAGCC	2220
GGCTCTTGGG GCGGTTTACA AGATTGTTGC AATCGAAGAT GAAACTGGTC AGATGCGCAA	2280
TACGATTAAG CTGTCTAATA ATGCTGAAAA AGTTTCTACG CCAGGTAAGA AGCAGGTGTG	2340
GCGCATTACC AGTCGTGAAA AAGGCAAGTC AGAAGGCGAC TATATCACTT ATGATGGTGT	2400
GGATATTAGC GACATGACAG AAATCAAGAT GTTCCATCCG ACCTATACAT ACATCAAGAA	2460
GACGGTTCGT AATTTTGATG CCGTTCCTCT CTGGTGGAT ATCTTCAAAG AAGGAATATT	2520
AGTTTACAAC TTGCCTAGTT TGA CTGACAT TCAGGATTAT GCCCGTAAGG AATTTGACAA	2580
GTTGTGGGAT GAGTATAAGC GTGTGCTCAA TCCGCAGCAC TATCCAGTGG ATTTGGCGCG	2640
TGATGTATGG CAAGATAAGA TGGACTTGAT TGATAAGATG CGCAAGGAAG CCCTTGGTGA	2700
AGGAGAAGAA GAATGAGTTT GCAAGAAAAG ATTATCCAAG AGCTGGGTGT CAAACCAGTG	2760
ATTGATGCCC AGGAAGAAAT CCGTCGTTCT ATTGATTCTT TAAAAAGATA TCTGAAAAAA	2820
CATCCCTTCC TAAAAACCTT TGTACTAGGG ATTTCTGGGG GACAAGACTC AACCTTGGCA	2880
GGACGTTTGG CGCAATTAGC TATGGAAGAA CTGCGAGCTG AAACGGGAGA CGATAGCTAC	2940
AAATTTATCG CTGTCCGCCT GCCATACGGA GTGCAAGCTG ATGAAGCAGA TGCTCAAAAA	3000
GCCCTAGCCT TCATCCAGCC AGATGTCAGC TTGGTTGTGA ATATCAAGGA ATCAGCTGAT	3060
GCCATGACAG CTGCAGTTGA AGCGACAGGT AGTCCTGTTT CAGACTTCAA CAAGGGGAAT	3120
ATCAAGGCAC GTTGCCGTAT GATTGCTCAG TATGCCCTTG CTGGTTCCCA TAGCGGAGCG	3180
GTCATTGGAA CAGACCACGC CGCGGAAAAT ATCACAGGTT TCTTTACCAA GTTTGGTGAC	3240
GGCGGTGCGG ATATTCTCCC TCTTTACCGC CTCAATAAAC GCCAAGGAAA ACAGCTCTTG	3300
CAGAACTTG GCGCAGAGCC AGCCCTTTAT GAAAAAATCC CAACGGCAGA CCTAGAAGAA	3360
GATAAACCAG GCCTAGCTGA CGAAGTCGCA CTTGGAGTCA CCTACGCAGA GATTGACGAC	3420
TACCTAGAAG GCAAAACAAT CAGCCCAGAA GCTCAAGCGA CCATTGAAAA CTGGTGGCAC	3480
AAAGGCCAAC ACAACGCCA CTTACCCATC ACCGTATTTG ATGACTTTTG GGAGTAAAAA	3540
GGTCCGGGGG ACCTTTT TAG CTTCTTGCCC TGAAATTAAA AAGCAAGAAA AACCTCCACT	3600

918

GGAGGTTTT	AGCCTCTCAT	CTTGAAATAA	GAAAGTGAGA	GAAGGTCTGG	GGGATCTTGA	3660
ACCCCGAGTT	TAGAAATAAG	AAAATGAGGC	AGATTTCAGTA	ACTCGAAGAG	TTCGATTTCA	3720
TCGTCTTACC	CCTGCAACGA	TGACTAGGTT	TGAAAAAGCT	TGCTAGAGCG	CATTTCAAAC	3780
CAGGCAGCAA	CTGCGTCAAG	AAATTAGAAG	ACAAACTCGT	TTTCTAGCTG	TACTGAGTT	3840
GAGCCTTTTT	ACTACGAGTA	TAGAAATAAG	GAAGTGAGGT	AGCATCATGA	AATCTATCGG	3900
TACGCAAATA	TTACAGACAG	AACGTTTGAT	TTTAAGAAGA	TTTGTGGAGA	GTGATGCAGA	3960
AGCCATGTTT	CAAAATTGGG	CTTCATCCGC	TGAGAATCTG	ACCTATGTTA	CCTGGGATCC	4020
CCATCCTGAT	GTCGAAATCA	CTCGAAACTC	GATTTGCAAT	TGGGTTGCTT	CCTATACTAA	4080
TCTCAACTAT	TATAAATGGG	CCATTGTGCT	AAAAGAAAAC	CCAGAGCAAG	TAATAGGAGA	4140
TATCAGCATT	GTTAAGATAG	ACGAGGCTGA	TTTAAGCTGT	GAAATTGGCT	ATGTGTTAGG	4200
CAAGGCTTAC	TGGGGAAATG	GTATGATGAC	AGAGACTTTG	AAAGCTATCT	TGGACTTTTG	4260
TTTTACTCAA	GCAGGTTTTC	AAAAGGTCAG	AGCACGTTAT	GCCAGTCTCA	ACCCAGCTTC	4320
AGGTCGTGTC	ATGGAAAAGG	CTGGAATGTC	CTATCTACAA	ACCATTGTTA	ATGGTGTTAGA	4380
GAGAAAAGGC	TATCTTGCGG	ATCTTATTTA	TTATGGTATA	AGTAGGGAAG	AATGTTGAAT	4440
TCTATTTTCT	GTTTCTATCG	AAGTCAACTA	TTTATTGTAA	ATATAATAAT	TAGCATTCCA	4500
AGTTTATTTG	AAACTTTAAA	ATAGCATATT	GATTAGTACA	AGACAGATGT	TCTAGTTCCT	4560
TCTTTAATCT	GGTTTAGTGT	TAGTTAAAAA	ATCGCTTTAA	GCTTGTAAC	AAGAGGGAGC	4620
TAATCGACTA	GATTCTCCAG	CCGAACAGGT	GGTAATGTAC	TTTTTATAGT	GTAATCCTAG	4680
CTGTTGTAA	ATTTAAAATA	GAATCCTCTA	TCGAGTTAGG	GAATTAAATT	CAACCAATTT	4740
TATTCATGTT	TTTTCTATCA	AATTATCTAA	TATTAAAATA	GTCTCATTCT	GATGAGAAAA	4800
CTATTCCCAA	ATCATTCTAA	CCTCTCTCAA	CTAGATGTAA	CTTACAAAAC	CCCTGACCTC	4860
ATGAGCCACT	TTCTTCCTCC	TCATGAGGTC	AGTTTTACTT	TCTGCTGTTT	CAGTATCGTT	4920
TTCTCTCGCT	AGATTTCCTC	AAAAGGGCAG	ACTCTCCCT	TGGTGCGTCA	CACGATTTTT	4980
TCATCTCGAC	TGTTCTTTAA	TGCATCATTA	ACGACGCTTT	TCTTCTAGGT	GGTTCATAAG	5040
GAACAGGAAG	ATTCAGGTTG	ACTTTTCTAA	TCCTAGAATA	AAGTGCTGAA	AACAATTCGG	5100
AATAGGCATA	GAGACTAGAC	AATTTGAGGA	GCTGCTTGCG	TCCTGTTTCA	ACACATTTTC	5160
CCACCACGTG	AAGAAAAAGA	TGGCGGAAGC	GTTTGATTGT	TAAAGTTTGG	AAGTCACCTC	5220
CAGCTAGATG	TTTGAGAAAA	AGATAGAGAT	TGTAGGCGAT	ACAGCTCATC	ATCATACGAA	5280
CTTCGTTTTT	GATTAAGGTT	GAAC				5305

(2) INFORMATION FOR SEQ ID NO: 136:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3964 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

TGGCAGCTCG TCGTCGTAAG GGACGCAAAG TTTTGGCTGC ATAATCCAAA CGAATTCTAT	60
CAAAAATCAG TAGGAAGCTG AGTCTACTGA TTTTATTTT TGTAAAAAG TTCAGTAGAT	120
GCAAATGGAT TCGGAAGCGA TGTACAGTA GATTGAACT AGAATAGTAC ACCTCTGTTT	180
CTAAACATT GTTAGAAATC GATTGACTG TCCTGATCGA TTTGCTCTGT TATTATTTTA	240
TTTACTATA AAGTTGAAGT AGGTGGAGAT GGTACAGCAA CAATCGTCTT TAAAGATGGT	300
TCAGCTATTA CAATTCCAGG AAATCAATTG GTAGCACAAG ATCCAAAAGC ACAAGATAGC	360
ACTAACTGA CTGCTGAAAA ATCAACTGTT AAAGCACCTG CTCAAAGAGT AGATGTAAAA	420
GATATAACTC ATTTAACAGA TGAAGAAAA GTTAAGGTTG CTATTTTACA AGCAAATGGT	480
TCAGCATTAG ACGGAGCGAC AATCAATGTA GCTGGAGATG GTACAGCAAC AATCACATTC	540
CCAGATGGTT CAGTAGTGAC GATTCTAGGA AAAGATACAG TTCAACAATC TGCGAAAGGT	600
GAATCTGTAA CTCAAGAAGC TACACCAGAG TATAAGCTAG AAAATACACC AGGTGGAGAT	660
AAGGGAGGCA ATACTGGAAG CTCAGATGCT AATGCGAATG AAGGCGGTGG TAGCCAGGCG	720
GGTGGATCAG CTCACACAGG TTCACAAAAC TCAGCTCAAT CACAAGCTTC TAAGCAATTA	780
GCTACTGAAA AAGAATCAGC TAAAAATGCC ATTGAAAAAG CAGCCAAGGA CAAGCAGGAT	840
GAAATCAAAG GCGCACCGCT TTCTGATAAA GAAAAAGCAG AACTTTTAGC AAGAGTGGAA	900
GCAGAAAAAC AAGCAGCTCT CAAAGAGATT GAAAATGCCA AAATATGGA AGATGTGAAG	960
GAAGCAGAAA CGATTGGAGT GCAAGCCATT GCCATGGTTA CAGTTCCTAA GAGACCACTG	1020
GCTCCTAATG CTGCTCCTAA GACAACAAGT GCACCGCAAG CAACTGCAGG AACAATGCAA	1080
GATGTTACCT ACCAGTCACC TGCTGGCAAA CAATTACCTA ACACAGGTTC AGCATCAAGT	1140
GCAGCACTTG CTAGTCTTG TCTAGTGGTG GCAACAAGTG GTTTTGCTTT GCTAGGAAGA	1200
AAGACTAGAC GTAGAAAATA GAACAGCTAG AAAATTCTAT TCTCTACTTA AAGTTAGATT	1260
ATAAGGGGGA TTTTGAGAAG TCATCAATCC TAGTGATGGG TGAGAAAAGT GAGAACCCTA	1320
GATAATCACA TACTTTAGCT GAATAGGAAT ATTCTATCAA TGAGCCAAT CTCTTCTGTC	1380
TCTAACTGTG GAATAGGAGA TGGCAATAT CGGATAGAAA AGATAGCAGA ATAGCTCTCT	1440

920

ATTGAAGAGA GGAGGGGAAA CCGAAAAATT AGGTGCCCCCT CCTCTTTTTT GGTATAATAG	1500
AAGATAGAAA ACGAGGTTAG AAGAGATGAT TTTTGATACA CATAACACT TGAATGTAGA	1560
AGAATTTGCA GGTCGTGAGG CAGAAGAAAT TGCCTTGGCT GCTGAGATGG GTGTGACACA	1620
GATGAATATT GTTGCTTTTG ATAAACCGAC GATTGAGCAT GCCTTGGAGT TGGTAGATGA	1680
GTATGAGCAG CTCTATGCCA CTATTGGTTG GCATCCTACA GAAGCTGGTA CTTATACAGA	1740
GGAAGTTGAG GCTTACTTGT TGGATAAGTT AAAACATTCC AAGGTTGTGG CTTTAGGTGA	1800
AATTGGCTTA GATTACCATT GGATGACAGC GCCCAAAGAG GTGCAGGAGC AGGTTTTTCG	1860
CCGTCAGATT CAGCTATCTA AGGACTTGGA TTTGCCTTTT GTTGTCCATA CCCGTGATGC	1920
GCTGGAAGAT ACCTATGAGA TTATCAAGAG TGAGGGCGTT GGTCTCTGTG GTGGTATCAT	1980
GCATTCAATT TCAGGGACGC TTGAGTGGGC AGAGAAGTTT GTGGATCTTG GTATGACCAT	2040
TTCTTCTCA GGAGTGGTGA CTTTAAAGAA GGCAACTGAC CTCCAAGAAG CAGCTAAAGA	2100
GTTACCTTTG GACAAGATGT TGGTGGAAAC AGATGCGCCT TACTTAGCAC CTGTACCCAA	2160
GCGTGGTCGT GAAAATAAAA CAGCCTATAC TCGCTATGTG GTCGACTTTA TCGCTGACTT	2220
GCGTGGTATG ACGACAGAAG AGCTGGCGGT AGCAACGACT GCAAATGCAG AACGAATTTT	2280
TGGACTGGAC AGCAAGTAAT GAAAGAGAAA ATTTCTCAAG TTATCGTGGT TGAAGGGCGT	2340
GATGATACGG TCAATCTCAA ACGTTATTTT GATGTGGAGA CCTATGAGAC TCGAGGTTCT	2400
GCCATCAATG CTCAGGATAT AGAGCGGATT CAGCGCCTGC ACCAACGTCA TGGAGTCATT	2460
GTCTTTACAG ACCCAGATTT TAATGGGGAA CGGATTCGGC GCATGATCAT GATGGTCATT	2520
CCAACAGTTC AGCATGCCTT TCTCAAGCGA GATGAAGCTG TTCCCAAGTC CAAGACCAAG	2580
GGGCGTTCTC TGGGAATTGA GCATGCCAGC TATGAAGACC TGA AACCGC TCTAGCTCAA	2640
GTGACAGAAC AATTTGAACA TGAGAGTCAG TTTGACATTA GTCGTAGCGA TTTGATTCGC	2700
CTTGCTTTTC TAGCAGGGGC AGACAGCCGT AAGCGTAGAG AATATCTCGG AGAGACTCTC	2760
CGAATCGGCT ATTCCAACGG CAAGCAACTC CTCAAACGCC TAGAGTTGTT TGGGGTACT	2820
TTGGCAGAAG TGAAGAAGC TATGAAATCT TATGAGTAGG AAAGATGTAG CCGTTACAA	2880
TTTTTAAGTT TCACAGTATT TTTCGAAGCA GGTAGAAGAG GAGGCGTCTG ATGTTAATTG	2940
GTCAAAAAAT TAAAGAGATT CGGATAGAAA AAGGAATTAG TCGTCCAGAT TTTGTGGAG	3000
ATGAGCAAGA ACTGACAGTT CGTCAACTGT CGCGAATTGA AAGTGGAGCT TCGCAACCGA	3060
GTTTGCCCAA GTTAGACTAT ATTGCTCGCC GGCTAGGAGT TCCAGTTTAT AGCCTTATGC	3120
CGGATTTTTC AGCTCTTCCT TCTGCTTATT TAGAATTGAA ATACCAGATT TTACGTGAAC	3180
CAATCTATGG TAAAGAAGAG GAGTACGATA AGAAGGAAGC GTGTTTGGAA GAGATTTATA	3240



921

AAACATACTT TGATAATCTT CCTAAAGAAG AACAAPTAGC ATGTGAAGTA TTGCAGGCGT	3300
GTTTGGATAC TTCTAGAACT AGAAGGCCTG AATATGCAGA GTTAATACTT GAGGAACATA	3360
TGCCTCAGAT TATAGAAAAA GAAGCTTATT CAATAAATGA TATGTTGTTG ATTCGTTTGT	3420
TTTTTTATCA AATGCTCATT AGAAAAGATC TTGCCAAATT TATAAATCAA ATCGAAAAGC	3480
TAATGCTCTT TCTTTTGGA CAGAAGAAGG TAACTCAAAT AGAGAATTAC TTTATAATTA	3540
GAGATACTCT TATTTAGGA ATGTGTTGTC TTGAAAAGGT AGGAGTAACT GATTGTTTTA	3600
ATGATTATCT ATCGTGTTTA CAAGAAATTA TGGATAAAAC TCAAGATTAT CAAAAGAAAC	3660
CTCTTGATTT TATGTTTTTG TGGAAGCAAG CATTAAAGAGA AGAAAGAGAT TTTAGTTTAC	3720
CTGAATCATT TTATCAGTCT TCTAAACAT TTGCGCAGCT AATTGGAGAT GAATTTCTAG	3780
TAAAGAAATT GACAGAGGAA TGGCAAGAGG ATGTCAAAAA ATATTTATAA ACATAGTGAA	3840
TCAGTGACAA AGATGTCCTT GTCCTCGTAT CAAAACAGTT CTAAAGTTTC TCTTTAGGGA	3900
TGTTTTTTTA GATATAAGCT AAAAATGACA CGAAATGGTT AGATTTTAAG GACATTGATG	3960
TCCG	3964

(2) INFORMATION FOR SEQ ID NO: 137:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12666 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

TGAGACCGTT ATTTGTATTA GGGAAATGGG TATCTATTTT TAATGCTGTG GGGATTTTGA	60
TTGTTTCTAT TATTCAAACC AAAAGCTTGT CAGGTATTGG AGCAGGATTG TTTAATCTAT	120
ATAACATTTT ATCTTATATA GGTGATTTAG TTAGTTTCAC TCGATTGATG GCATTAGGAT	180
TATCTGGAGC AAGTATAGCA TCAGCTTTCA ATTTAATTGT TGGTTTGTTC CCGGGAATAT	240
TGGCTAAACT GACAATTGGA TTAGTATTAT TCATTCTTTT ACATGCGATC AATATTTTTC	300
TATCGTTACT ATCAGGATAT GTTCATGGAG CACGCTGAT ATTTGTTGAA TTTTTTGGTA	360
AGTTTATGA GGTGAGGGA AAACCATTTT AACCTTTGAA GGCTTCTGAG AAATATATTA	420
AGGTTATTAC AAAGAATTAA TGGAGGATAT ATATAATGGA ACATTTAGCA ACTTATTTT	480
CAACCTATGG AGGAGCTTTC TTCGCTGCAT TGGGAATTGT ATTGGCGGTT GGATTAAGCG	540
GTATGGGGTC TGCTTATGGA GTTGTAAGG CTGGGCAATC TGCCGCAGCT TTAATGAAAG	600

922

AACAGCCTGA AAAGTTTGCC TCAGCTTTGA TATTGCAATT ATTGCCCGGA ACACAAGGAT	660
TATATGGTTT TGTATTGGA ATTTTAATTT GGTGCAATT AACTCCAGAA CTTCTTTAG	720
AAAAAGGCGT TGCTTATTTT TTTGTAGCTC TTCCAATTGC TATTGTAGGA TACTTTTCAG	780
CTAAGCATCA AGGAAATGTA GCAGTAGCGG GAATGCAAAT CTTGGCTAAA AGACCAAAAG	840
AATTCATGAA GGGAGCAATT TTAGCTGCCA TGGTAGAAAC CTATGCAATT CTTGCTTTTG	900
TCGTATCATT CATTGTGACC CTTCTGTAT AAGAAATAAA TTTGCAATTC AAAGGAGGTG	960
TCTAAATGAG CAATTTAGAA AACTTACGAG AGTCTGTTAT TGAACAAGCT CATGAAAAAG	1020
GGCGTATGAA ATTATTGGAT TCCAAAAAGA AGATTGATGA TGAATTTGAA ATGCAAAAGT	1080
CGCTCATTAT AAAGAAAAAA GAAGCTGAAC ATGAACGAAA GTTAAAAGAA TTGCAACAGA	1140
AATATCAAAT AATTTTCAA CAATTAAAA ATAAGGAACG CCAATCAACG TTAGTATCAA	1200
AACAGAAAAT ATTTAAAGAA CTTTTCAT CTGCTTTACT AGAAATGGAA TCTTGGAGTG	1260
CAGATAAAGA AATGGAGTTC ATCTATCGAA TTCTGGAACG ATATTCACAA CAAGAGGTCA	1320
TAGTAACCTT TGGGGAACGG ACTTTAGCTA AATCAATTT GGAACAATTA GAGAAATTGA	1380
AATTCTCTTT TCCAAATTAT TTATTTAGTG AACAACTAT CTCAAATGAA TCAGGCTTAC	1440
TTATTTCAAT AGGTAAAATT GATGATAACT ATTTGTATAA AACATTAATT GGATCGATTT	1500
CTAAGGAAGA AAGTTCAAGT ATCGCAAATC AAATTTTAT CAATTAAGGA TGAATTTGGT	1560
TAATCCTTCT TAGAAATTG GAGTATTCCA ATAAATTAG AAAGGTATTT TATGGATACT	1620
AATCTTTTTT CAAAAATAA TACGACGATT TCGGTAAAAG AAAACGATTT TATTACAGAA	1680
GAAAAATTTC AAAAAATTAT ACAATCCAAA GATACGGAGA CATTGGCATT TATCTTAGAA	1740
TCAACTCCCT ATCATTATC GATTGACATC TTAGAAGATC CTAGTCAGAC AGAGATTTCTG	1800
CTAATGACAA AATTAGTCAA TGATTATAGA TGGGCCTATG CTGAAAGTCC GTCTGATATA	1860
ATTGTGACTT TATTTGCTTT ACGATATGTT TATCATAATA TCAAAGTTTT ATTTAAATCT	1920
AAGGCGGCAA TTAAGAAAGA TTTTCTAAA TTATTAATTC CAATAGGGAT TTTTGATATA	1980
GAAAGTTTAA AACATTTAGT TTCTTCCTTA CATTGAGATA CACTTCCTGA TTTTATGGTT	2040
CGTGAAGTAG AATCAATTTG GAATGAGTAT GAACTTTTA ATAATATTCG TGTACTTGAT	2100
GTCGGAGCTG ATCTAGCATA TTTTAAACAT CTGAACTTT TATCTAATGA GTTAGATGAG	2160
GTAATGTCTC AGGTATTGT CGAAATGATT GACTTTTATA ATATTATTAC TGTAACCGT	2220
GGTTTATCTC AAAATAAGAG TCATGGGGAT ATTTTACAAT TACTTTCAGA TGAAGGAAGT	2280
ATTTCTGCTA AAGAATTTAT ATACATTGTA GAAATCAAG AAATATTTGT GTGGTTCAAT	2340
AAAATAAATC CAAGCTTAGA TTCAATCTTT TCAACTTATG AATTGAAGAT GCAGGACGCA	2400

923

ACAATTTTCAT CTTCTGAGTT AGAATTTTTA TGTGATTAC TATTGTATAA AACTTTAGAT	2460
CAAGGAAGGT ACAATGTAGA GGGGCCGTTA GTTCTTGCTA GATATTTATT GGGATGTGAG	2520
TTTGAAGTAA AGAATCTCAG AATGATCATA TCAGCTCTTC AAAATACAAT TCCCTTTGAA	2580
TCAATAAAAG AAAGGATACG CCCACATTAT GGAAGCTAAT AAGTATAAAA TTGGCATAAT	2640
TGGTAGCCGT GATATTATTT TACCATTTAG CATGATTGGG TTTGATATAT TTCCTGCCTA	2700
CCAAGAACAA GAAGCTATAA ATACACTAAG AAAATTAGCT CAATCTGATT ATGGTGTGAT	2760
TTATATCACG GAAGACATTG CTTCAATGAT ATTAGATACA ATTCGCCATT ATGATTCCCA	2820
AGTTGTGCCT GCTATTATTT TATTACCGAC TCATAAACAA GGTTTAAATT TAGGATTAAA	2880
ACGTATAGAG GATAATGTAG AGAAAGCAGT AGGACACAAT ATTTTATAAT AATGTACAAA	2940
ATTGTCTGTA ATATTATTCT ATAATTTTGT GACTTAGTAA GGAGAATAAC TTTGACTCAA	3000
GGGAAGATTA TAAAAGTATC GGGACCTCTA GTTATTGCAT CAGGTATGCA GGAGGCTAAT	3060
ATTCAAGATA TTTGCCGTGT AGGTAAGCTA GGGTTAATCG GTGAAATTAT TGAAATGAGA	3120
AGAGATCAGG CATCTATCCA AGTCTATGAA GAAACATCTG GTCTTGGTCC GGGAGAACCT	3180
GTGTGTACAA CTGGAGAACC TCTCTCGGTT GAATTAGGGC CAGGATTGAT TTCTCAAATG	3240
TTTGATGGCA TACAACGCCC ATTAGATCGA TTTAAATTGG CTACTCATAA TGATTTTCTA	3300
GTTCGTGGGG TAGAAGTTCC AAGTTTGGAT AGAGATATTA AGTGGCATTG TCATTCCACT	3360
ATAGCAATTG GTCAAAAAGT GAGTACGGGT GATATTCTTG GAACTGTCAA GGAAACCGAG	3420
GTAGTTAATC ATAAAATTAT GGTTCCTTAT GGAGTATCTG GAGAAGTCGT TTCTATTGCA	3480
TCTGGCGATT TTACAATTGA TGAAGTTGTA TATGAAATAA AAAAATTGGA CGGTAGTTTC	3540
TATAAAGGAA CGCTTATGCA AAAATGGCCT GTCCGCAAGG CGCGTCCTGT TTCTAAACGT	3600
TTAATTCAG AAGAACCATT AATCACAGGT CAACGAGTTA TTGATGCATT CTTTCCAGTA	3660
ACCAAAGGGG GAGCTGCAGC AGTTCCTGGA CCGTTTGGAG CAGGAAAGAC AGTTGTACAA	3720
CACCAAGTAG CTAAATTTGC CAATGTTGAT ATTGTTATTT ATGTCGGTTG TGGAGAACGT	3780
GGAAATGAAA TGACGGATGT ACTGAATGAG TTTCTGAGT TGATTGACCC TAATACCGGA	3840
CAATCAATTA TGCAACGGAC AGTTCTGATT GCTAATACTT CAAATATGCC TGTGCTGCT	3900
CGTGAGGCTT CAATTTATAC AGGAATTACC ATGGCTGACT ATTTTCGTGA TATGGGCTAC	3960
TCTGTCGCCA TTATGGCTGA TTCAACTTCA CGTTGGGCAG AAGCGCTACG TGAAATGTCA	4020
GGACGTCTAG AAGAAATGCC TGGTGATGAG GGTATCCTG CTTATCTGGG AAGTCGTATC	4080
GCTGAATATT ATGAAAGAGC AGGACGTTCT CAGGTTCTAG GGCTTCCAGA ACGTGAAGGA	4140

924

ACGATTACTG	CTATTGGAGC	TGTATCGCCA	CCTGGTGGAG	ATATTTTCAGA	ACCAGTTACT	4200
CAAAACACTT	TACGGATTGT	GAAAGTTTTT	TGGGGGCTTG	ATGCTCCGTT	GGCACAGCGA	4260
CGTCATTTTC	CTGCAATTAA	CTGGCTTACA	TCTTATTCAC	TATATAAAGA	CAGTGTGGGC	4320
ACTTATATAG	ATGGTAAAGA	GAAGACAGAT	TGGAATAGTA	AAATAACTCG	TGCGATGAAC	4380
TACTTACAAC	GGGAATCTAG	TTTAGAGGAA	ATTGTTTCGTC	TTGTTGGAAT	TGATTCTCTG	4440
TCTGATAATG	AACGACTAAC	GATGGAAATT	GCTAAACAAA	TTGAGAAGA	TTATTTGCAA	4500
CAGAACGCTT	TTGATTCGGT	AGATACATTC	ACTTCGTTTG	CAAAACAAGA	AGCAATGCTA	4560
AGTAATATTC	TCACTTTTGC	TGATCAGGCA	AATCATGCTT	TAGAGTTGGG	TTCTTACTTT	4620
ACAGAGATTA	TGGAAGGTAC	CGTGGCAGTT	CGAGACCGTA	TGGCGAGAAG	TAAATATGTT	4680
TCAGAAGATA	GATTAGATGA	AATCAAAATT	ATATCAAAATG	AGATTACACA	TCAAAATTCAT	4740
TTGATATTAG	AAACAGGAGG	TCTATAAATG	AGTGTATATA	AAGAATACAG	AACTGCTAGT	4800
GAAGTTGTTG	GGCCTCTTAT	GATTGTTGAA	CAAGTAAATA	ATGTGTCTTA	CAATGAGTTA	4860
GTTGAAATTC	AAC TTCATAA	TGGAGAAATT	CGTCGTGGAC	AAGTTTTAGA	GATCCACGAA	4920
GATAAAGCAA	TGGTTCAGCT	TTTTGAAGGA	TCTAGTGGAA	TAAATTTAGA	AAAGTCTAAA	4980
ATTCGTTTTG	CTGGTCATGC	ATTAGAATTG	GCTGTATCTG	AGGATATGGT	TGGTCGTATT	5040
TTTAATGGGA	TGGGAAAACC	AATTGATGGT	GGACCAGATT	TAATCCAGA	GAAATATTTA	5100
GATATTGATG	GTCAAGCTAT	TAATCCTGTA	TCTAGAGATT	ATCCAGATGA	ATTTATTCAG	5160
ACAGGGATCT	CCTCTATTGA	TCATTTGAAT	ACTCTTGATC	GTGGTCAAAA	ATTACCAGTA	5220
TTTTCAGGTT	CGGGCTTACC	TCATAATGAA	TTAGCTGCTC	AGATAGCAAG	ACAAGCGACT	5280
GTTTTAAATT	CTGATGAAAA	TTTTGCGGTT	GTATTTGCAG	CAATGGGTAT	TACTTTTGAA	5340
GAAGCTGAGT	TTTTTATGGA	AGAACTCAGA	AAAACAGGAG	CGATCGATCG	TTCCGGTTTTA	5400
TTTATGAACT	TGGCAAATGA	TCCTGCAATT	GAGCGTATTG	CAACTCCCCG	CATTGCTTTA	5460
ACTGCGGCAG	AGTATCTAGC	TTTTGAAAAA	GATATGCACG	TTCTAGTTAT	CATGACGGAT	5520
ATGACTAACT	ATTGTGAAGC	GTTACGTGAA	GTCTCGGCAG	CTCGCCGTGA	AGTTCCAGGG	5580
AGACGAGGCT	ATCCGGGATA	TTTATATACA	AATTTATCAA	CTCTATACGA	AAGGGCTGGT	5640
CGCTTAGTTG	GTAAAAAAGG	TTCCGTGACA	CAGATTCCCTA	TTTAAACAAT	GCCAGAAGAT	5700
GACATAACAC	ATCCAATTCC	TGATTTAACT	GGATACATTA	CTGAAGGGCA	AATTTATTTG	5760
TCGCATGAGT	TGTATAATCA	AGGTTATCGT	CCACCAATCA	ATGTTTTACC	TTCTCTCTCT	5820
CGATTAAAAG	ATAAGGGATC	TGGAGAAGGT	AAAACCTCGT	GAGATCATGC	TCCAACATATG	5880
AATCAACTGT	TTGCAGCCTA	TGCCCAAGGG	AAAAAGGTTG	AAGAGTTAGC	AGTAGTATTA	5940

925

GGAGAATCGG CTTTATCTGA TGTAGATAAA TTGTATGTGA GGTTTACAAA CCGTTTGTAA	6000
GAAGAGTACA TAAACCAAGG ATTTTATAAA AATCGAAATA TAGAAGATAC GTTGAATCTT	6060
GGGTGGGAAT TACTATCAAT TCTTCCTAGA ACAGAGTTAA AACGTATCAA AGATGATTTG	6120
CTTGATAAAT ACTTACCTTT GGTAGAAGTT TAATCCGGAA ATGGAGTGAT TATCTATGGT	6180
ACGTTTGAAT GTAAAACCAA CTCGTATGGA ATTGAATAAC TTAAAGGAAC GTTTGACAAC	6240
AGCTGAACGT GGACATAAGT TATTAAAGGA TAAAAGAGAT GAATTGATGA GGCGATTTAT	6300
TTCTTTGATT CGTGAGAATA ATCAACTTCG GAAAGAAGTG GAAAGTTATC TAATTGATAA	6360
TCTAAATCC TTTGCAGTTG CTAAATCATT AAAGAATTCT CAAATGGTGG AGGAATTATT	6420
TTCAATTCCA TCGAAAGAAA TTGAATTATT TGTGAGAAA GAAAATATCA TGAGTGTAAC	6480
AGTTCCTAGA ATGCATATGA ATATTACTTC TCAAAATGAG AACAGTGAAT ACAGCTATTT	6540
ATCTTCTAAT AGTGAATGG ATGATGTATT TGCTACAATG AATAGTTTAA TTTATAAATT	6600
ACTAAGACTG GCAGAACTTG AAAAAACGTG TCAGTTAATG GCTGATGAAA TAGAAAAAAC	6660
ACGTAGACGT GTAAATGGTT TAGAATACTC GATTATTCCA AACTTGTCGG AACTATTCA	6720
TTATATAGAA TTGAACTAG AGGAGGCAGA AAGAGCCAAT TTAGTTCGTA TTATGAAAGT	6780
GAAGTAGATC CTTTATTTAG ATTATTAATT AGATGAACAA ATATCAGCTT GGATAAGGCT	6840
TTAAGCCTTT CTAAGCTTTT TTTATTGACA GTATCAGGAT ATCTTTTCA AAATTTTGGT	6900
TTGTTAGATA ATGAAATGT TTCTACTAAT CTAGATTTAG GATTAGTAAA TCGTAAATGT	6960
AATTATATAG AAAGTAAGCG CGTCATAACA AGGTATCTAT CATTCATGGA GCTCCTCCTG	7020
TATACTATTA GTAAAGTAAA ACTATTGGAG GATATTTTAA TGCCACAACC TATTGTTCTT	7080
GTAGAGATTC CACAATCTCG TCGTTTTGAT TCTAAAAAGA GAAATGATAT TCTGCTTAAA	7140
ATTCGTATTG GCAAGCTTGA AGTAAGTTTT TTTCAATCTC TCAATCTCGA AATGGTAGAA	7200
CAGCTTTTGG ATAAGGTGTT GCTCTATGAC AATTCATCTA TCTAGCCTAG GGGAGGTCTA	7260
TCTCGTGTGT GGGAAAAC TG ATATGAGACA AGGAATCGAT TCACTGGCTT ATCTGGTTAA	7320
AACCCACTTT GAATTGGATC CTTTCTCCGG TCAAGTCTTT CTCTTTTGTG GTGGACGTAA	7380
AGACCGCTTT AAAGTCCTTT ACTGGGATGG TCAAGGATTT TGGCTACTAT ATAAACGCTT	7440
TGAGAACGGC AGATTGATTT GGCTAAGTAC AGAAAAGGAT GTCAAAGCTC TCACACCAGA	7500
ACAAGTAGAC TGGCTTATGA AGGGCTTTTC TATCACTCCA AAAATATAGT AGATTGAAAC	7560
TAGAATAGTA CACCTCTGCT TCTAAAACAT TGTTAGAAAT CGATTTTACT GTCCTGATCG	7620
ATTTGTCCTG TTCTTATTTT ATTTTACTAT AAATCCATCA GAAAGTCGTG ATTTCTATTG	7680

926

AAATGAGGAC	TTTCTTTT	TACTCATCTG	CTTTCAAAA	GCATTCTAGT	CCATCTCCGA	7740
TTAACGATGG	ACTTTATCAC	CTCCTTCTCC	AGTCCTTGTA	TAACATCTTG	GAGTTGATTG	7800
ATGACATCTT	CCAAAGTTTA	AAAGGCTTTA	TTCTTAAATC	CACGTTTACG	AATCTCTTTC	7860
CACACTTGTT	CAATGGGGTT	CATCTCTGGT	GTGTATGGAG	GAATAAATGC	AAAGCCAATA	7920
TTAGTCGGAA	TCTTTAAGGT	ACTTGATTTA	TGCCATATAG	CATTGTCCAT	AACGAGTAAA	7980
AGATAATCAT	CTGATAAGC	TTGTGAAATC	TCCTATTCTT	AAAGCCCCTT	TAGCGCATAA	8040
CTTTGGCTCA	GCTTCTATTA	TCGCTCACAC	CATCCATCAG	AAGTTTAATC	TGAAGGTACC	8100
CAATTATCGC	CAAGAAGAAG	ATTGGGCTAG	GATGGGTTTA	CCAATCACAC	GTAAGGAAAT	8160
CTCTAATTGG	CATATCAAGG	CGAGTCAATA	CTATTTGGAG	CCCCTTTATA	ACCTCTTGCG	8220
AGAGAGACTA	TTGACTCAGC	CCTTACTTCA	TGCGGATGAA	ACTTCTTATA	GGGTGCTAGA	8280
GAGTGATAGT	CAGCTGACTT	ACTATTGGAC	TTTTTTGTCA	GGTAAAGCAG	AGAAACAAGG	8340
GATTACGCTT	TACCACCATG	ATCAGTGTG	AAGTGGTTCA	GTAGTACAAG	AATTCCTAGG	8400
AGATTATTCT	GGCTATGTGC	ATTGTGATAT	TTTGCGGCAG	TAACCTTAGGA	CTTTAGTCCT	8460
CTAGTTCTGC	CTATGCGATA	GCAGTCCAAG	GTTTAGGAGC	AAGGCGACGC	TAAGCTTGGT	8520
AAACTTCGAA	CCGCTCGTCT	GCTTATCGTC	AACTGGAAGA	AGCTGAACTT	GTTGGATGTT	8580
GGGCGCATGT	GAGAAGGAAG	TTTTTTGAAG	CGCCCCCCCA	AGCAAGCGGA	TAAATCATCC	8640
TTAGGAGCTA	AAGGTTTAGC	TTATTGTGAT	CAGTTATTTT	CCTTGGAAG	AGACTGGGAG	8700
GCTTTGCCAG	CTGATGAACG	ACTACAGAAA	CGTCAAGAAC	ATCTCCAGCC	CTTAATGGAA	8760
GACTTCTTTG	CTTAGTGCCG	GCGTCAGTCA	GTTTTAGCAG	GTTCAAAACT	AGGAAGGGCA	8820
ATTGAATACA	GCCTCAAGTA	TGAAGAAACC	TTTAAGACCA	TTTTGAAAGA	CGGACATCTG	8880
GTCCCTTTCCA	ATAATCTAGC	TGAACGCGCC	ATTAAATCAT	TGGTTATGGG	ACGGAGTAAA	8940
AGAGTCCAGT	GGACTCTTTT	AGCCTAAGCT	CAGTTTAAAA	AAGCGAGGGT	GGTTATTTTC	9000
TCAAAGTTTT	GAAGGAGCTA	AAGCAAGAGC	TATTATTATG	AGTTTGTTGG	AAACAGCTAA	9060
ACGTCATCAA	TTAAATAGCG	AGAAATATCT	ATCCTATCTT	CTAGAATGTC	TTCCAAACGA	9120
GGAAACTCTC	GTAAACAAAG	AGGTTTTAGA	GGCTTATTTA	CCATGGACTA	AAGTTGTACA	9180
AGAAAAGTGC	AAATAAGAAA	TCTCCAGATT	AGGAACTATC	CGTGAGTTCT	CCAGTCTGGA	9240
GATTTTTCAA	TAGACTTCCT	GCGAAACAAA	ATATGGTATA	ATAGTTCTAT	GAATGATGAA	9300
GCAAGTAAAC	AACTAACCGA	TGCACGATTT	AAGCGTCTTG	TTGGTGTTC	ACGCACGACT	9360
TTTGAAGAGA	TGTTAGCTGT	ATTAAAAACA	GCTTATCAAC	TTAAACACGC	AAAAGGTGGA	9420
CGAAAACCTA	AATTAAGTCT	AGAAGACCTT	CTTATGGCCA	CTCTTCAATA	TGTGCGAGAA	9480

927

TATCGAACTT ATGAACAAAT TCGGCTGTT TTTGGTATTC ACGAAAGCAA CTTAATCCGT	9540
CGGAGCCAAT GGGTTGAAGT AACTCTTGTT CAAAGTGGTG TTACGATTTC AAGAACTCCT	9600
CTCAGTTCTG AGGACACGGT AATGATTGAT GCGACGGAAG TAAAAATCAA TCGCCCTAAA	9660
AAAAGAATTA GCGAATTATT CTGGTAAAAA GAAATTTTAC GCTATGAAGG CTCAAGCGAT	9720
TGTCACAAGT CAAGGGAGAA TTGTTTCTTT GGATATCACT GTGAACTATT GTCATGATAT	9780
GAAGTTGTTT AAAATGAGTC GCAGAAATAT CAGACAAGCT GGTAATCTT TGGCTGACAG	9840
TGGTTATCAA GGGCTCATGA AGATATATCC TCAAGCAGAA ACTTCACGTA AATCCAGCAA	9900
ACTCAAACCG CTAACAATTG AAGATAAAGT CTATAACCAT GCGCTATCTA AGGAGAGAAG	9960
CAAGGTTGAG AACATCTTTG CCAAAGTAAA AACGTTTAAA ATGATTTCAA CAACCTATCG	10020
AAATCATCTA AACGCTTCGG ATTACGAATG AATTTGATTG CTGGTATTAT CAATCATGAA	10080
CTAGGATTCT AGTTTTGCAG GAAGTCTATT ATCAAAAATA CCATCAAGAT TATATAAGAT	10140
TGATACAGGA AAAGTTTTAT TTGATGGTGT AAATATTAAT CAAATAGATA AAAAAATATT	10200
AAGTCAAAAT TTAGGAGTAG TTCCACAGGA TTCATTTTTA TTGAACCGAA GTATTCTTGA	10260
TAATATAACT TTAAAGCAGG AAGTTACTTC ACAAAGATA GAGGAAGTTT GTAAAGCAGT	10320
TCAAATCTAT GATGAAATCA TGCCTATGCC GATGAAATTT AATACTATCA TCTCAGAGAT	10380
GGGGTCAAAT ATTTACGGTG GGCAAAGGCA ACGGATAGCA CTGGCACGTG CATTAATAAA	10440
TAATCCTAGT ATTGTAATTT TAGATGAAGC AACTAGTGCA TTAGACACTA TTAATGAGGA	10500
AAGAATAACA AAGTATATAC AAAGTCAGGG CTGTACTCAA ATAATTGTAG CTCATAGATT	10560
GTCAACGATT AAGGATGCGG ATGTTATTTT TGTAATGAAA GGTGGTAAGA TTGTTGAGTC	10620
AGGAAATCAT AAGTACTTAA TGGATCTTGG TGGAGAGTAC TACAGCTTAT ATACAAAAG	10680
GAAATGAGGT GTAAAGAAAA TGAAGAAAGA AAATGAATAT GTAATTTTAA CAACAGCCTC	10740
ACTAGGGGTG ATGATTGGAA TAGTGTGTTG AATTTTTTTA GATTTTCCAG TTGAATATGG	10800
TATTTCTTTA GGCTTGTTGA ATGGAATAGT ATTGGGTTG CTGATTGTTT AAAAAACAA	10860
TAAGAATTAA GCATAATTTT TTGCTGTAAA CTAAGGAGTA GAGATGGCTA TAGTTGAAAT	10920
TATAAATCTA ACAAAGGCT TTAAAGATAT TGAAGTTATT CATAACACTT AAATAATAGA	10980
GCAACTACAG TAGTAGCTTA AAAACATGAT TAAATCGCTA TTCTTAGGAG TAGCGGTTTT	11040
TCTTTTGTG TAATACTCTT TGAAAATCTC TTCAAACCAC GTCAGCTTTG CTTTACCGTA	11100
CTCAAGTACA GCCTGCGGCT CGCTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATAAAA	11160
AGGGTCAAGT AAGTATAGTA AATTGAAATA AGATATGAAC AAATCGATTA GAAAAGTCAA	11220

928

ATTAATTTCT AGAAATATGT TAGAAATTGG TTTGAATTCC GCAATCAATT TGTTCAGTTT	11280
TTATTTTCATT TCATTTTATT TAATTAGATT TTCCAATTTT TTAATTCAG CTAAAAATCC	11340
CCAATCGTAG TGATTGAGGA TTGAGTAAAT AAATCTTAAA CAATACCTTG TGCAATCATG	11400
GCATTTGCTA CATTTCAAA GGCAGCAATG TTAGCTCCTG CAAGGTAGTC TTTATCAAGA	11460
CCGTATGTTT CTGAAGTCGT TTTAGCTGTG TTGAAGATGT TTGTCATGAT GTCTTTGAGA	11520
CGGCCATCAA CTTCTTCACG AGTCCATGAG AGGCGAAGAC TGTTTTGGCT CATTTCAAGA	11580
GCTGAAACGG CTACACCACC AGCGTTGGCA GCTTTTGCAG GTCCGTAGAA GATACCATTT	11640
TCTTTGTAAA CTTTGATGGC ATCAAGGTCG CTCGGCATGT TGGCACCTTC AGATACACAG	11700
ATAACGCCTT GAGCAACCAA ACGTTTAGCT GCTTCACCGT TGATTTCTGT TTGAGTGGCA	11760
CATGGAAGAG CAATGTCATA GTTCCAGCG TAAGTCCATA CAGTACCTTC GTGGTAGGTT	11820
GCAGTTGCTT TTTCAGCTGC ATACTCAGTC AAACGAGCAC GACGTTTTTC TTTAACATCA	11880
ACCAAAAGAT CGAAGTCGAT ACCATTTTCA TCGATGACAT AACCATTTGA GTCAGAAACA	11940
GAAATAACAG TTGCACCGAG TTCAGTTGCT TTTTGAAGAG CATATTGAGC AACGTTACCA	12000
GAACCTGAAA TAACGACTTT CTTACCAGCA AAGCTGTTAC CGTTAGCTTT GAGCATTTCT	12060
TCAGTATAGT AAACCAAACC GTAACCAGTT GCTTCTGGAC GAATCAAGCT ACCACCAAAT	12120
CCAAGAGGTT TACCAGTCAA GACACCAGCA TCAAATTGGT TAAGACGTTT GTATTGACCG	12180
TAAAGGTAAC CAATTTACG TCCACCAACA CCGATATCAC CAGCAGGTAC GTCAAGTGAT	12240
GGTCCGATGT GTTTTTGCAA TTCAGTCATG AAGCTTTGGC AGAAGCGCAT CACTTCAGCA	12300
TCTGTTTTAC CTTTAGGATC GAAGTCTGAT CCACCTTTAC CTCCACCGAT AGGAAGTCCA	12360
GTCAAGACAT TTTTAAAGAT TTGTTCAAAT CCGAGGAATT TCAAGATCCC TTGGTTTACA	12420
GTTGGGTGGA AACGAAGTCC ACCTTTGTAT GGTCCAACAG CTGAGTTGAA TTGAACACGG	12480
TAACCACGGT TTAATTGAAT TTTTCCATCA CGGTCAACCC AAGGAACACG GAAAGAAACC	12540
ACGCGCTCAG GCTCAGTAAT ACGTGCCAAG ATATTTTCTT CGATATACTC AGGGTGTTTT	12600
TCAAATACAG GTTCTAAAGT GTTGAAAAAT TCTTCAACAG CTTGGAGGAA TTCAGCCTCG	12660
TGCCCG	12666

(2) INFORMATION FOR SEQ ID NO: 138:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1083 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear



929

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

AGCAACTGTT GTGAACCAAT TCCGATAAAT TCCAAGAATT GGTTAATAGA GCCATTTTGA	60
CCAAAAATCC CGATAAAAGC ATAGGCTTTA AGGAGCAAAT TGATCCAGGT AGGAAGGATA	120
ATCAGCATGA GCCAGAGTTG ACGGTGTTTG AGACGGGTCA AAAAGAGGGC CGTCGGATAA	180
CTGATAAGCA GTGCCACAAA GGTCACAATG CCTGCATAAA GCACTGAGTT GAAACTCATT	240
TTAAGATAGG TCAAGTTTTG TGACGCAAAG TAAGATTGT AATTTTCTAA ACTGAACTGG	300
CCTTCGATGT TGAAAAAGGA TTGACCGAAA ATCAAGACCA AGGGTGCCAA TACAAAGAGC	360
GCAATCCAAA GCATGTAGGG TACTACAAAG AGTTTAGAGC TTGTTTTCTT CATCTCTTTC	420
CTCCTCGATT GCATTGATCA AACCTGCTTC TTGCTCTTCG ATTTCTACGT ACTCCTCAAT	480
ACGAGCATCG AACTCTTCTT CGGTTTCATT GAGACGCATG ATGTGGATGT CTCTGGTTC	540
AAAGTCCAGA CCGATTTCCT CACCCACGAT AGCCTTACGG GTTGAGTGGA TCATCCATTC	600
ATTTCCAAGT TCGTCATAGG CGATAATTC ATAATGAACT CCACGGAAAA GCTGGGTATC	660
GACCTTAACT TGGAGCTTGC CTTCTTCAGG AAGGTAATG CGCAAGTCCT CTGGACGAAT	720
AACGACCTCA ACAGGTTTCAT TTGGCTTCAT CCCACCATCA ACCGCTTCAA AGCGTTTGCC	780
GTTAAATTCG ACCAAGTAGT CCTCAATCAT GGTACCTGGC AAGATGTTTG ACTCCCGAT	840
AAAGGTGGCA ACAAAGTGGT TGATTGGCTC ATCGTAGATG TCCACAGGGG TTCCAGACTG	900
GACAATCTCG CCATCATTCA TAACGAAAAT CCAGTCACTC ATGGCAAGAG CTTCTTCCTG	960
ATCGTGAGTG ACAAAGACAA AGGTAATGCC CAATCGTTGT TGTAAATCAC GCAATTCGTA	1020
CTGCATGTCT GTTCTCAATT TCAAGTCCAG CGCTGATAAA GGCTCGTCCA ACAAGACCAC	1080
ACGGGGTTGG TTGATGATAG CACGGGCGAT GGCCACACGC TGACGTTGTC CTCCAGAAAG	1140
TTTGCGGATG GAACGTTTTT CATAACCTTC CAACTGAACC ATCTTGAGAA CTTCCGCTAC	1200
ACGCTGCTCG ATTTCTTTCT TATCAATTTT ACGCAAGCGA AGTGGAAGG CAACATTTTC	1260
AAACACATTC ATATGTGGGA ACAAGGCATA GGATTGGAAG ACGGTATGTA CGTCGCGCTT	1320
GTTGGTTGGA ATATCATTGA TACGAACACC GTCTAGCATG ATATCTCCTG TCGTCGCATC	1380
CAGTAAACCT GCAATAATGT TTAGGATAGT TGATTTCCCC GAACCAGATG CACCTAGAAG	1440
GGTGTAGAAT TTCCCTTCTT CCAACTCAA GTTGATGTCT TTGAGAACCT TGGTGTGCT	1500
GTCTTCAAAA ACTTTAGAGA CGTTTTTGAA TTCGATAATT GGCTTTTCA ATTGGCATAA	1560
ATTCCTTCTT TTTCATAGAT TAACCGATCG GGGCTCTGTC AGGTCCCCAC TACCTCTTGC	1620
AGGGAGTAAA ACCACCTGCA TACATCTTCG CTACCGATAG GCTTTCACCC AAGATCCGGA	1680

910

CTTCTCTTC AAGCGTAATA CCTGAGTGTT CCTTGACTTT TCGATAACC GATTGGATCA	1740
AGTCCTCGTA GTCTTTGGCC GTTCCATCTG CGACATTGAT CATAAATCCT GCATGCTTTT	1800
CTGACACTTC TACGCCACCG ATACGATAGC CTTTCAAGCC AGCTTCTGAA ATTAAGTGAC	1860
CTGCAAAATG CCCGACTGGA CGCTTAAAGA CCGAGCCACA AGATGGGTAT TCCAAAGGTT	1920
GCTTGAGTTC ACGTAGGTGC GTCAAGCGGT CCATTTCTCTG CTTGATAACC TGATGGGTTC	1980
CTGGAGCTAG GGCAAATTTA ACTGACAAGA CAACTGCACC AGACTCCTGA ATAGCTGAAT	2040
GACGGTAACC AAAAGCCAAG TCTTTAGCAG ACAGGGTTTC GATTTCTCCA TCCTTGGTCA	2100
AGACCTTACA AGACTGCAAG ATGTGAGCAA TCTCGCCACC ATAGGCACCC GCATTCATAA	2160
AGACAGCACC GCCAACGCTT CCTGGAATAC CACAAGCAAA CTCAAAGCCA GTTAAACTAT	2220
GACGGAGGGC AATGCGAGTT GTTTCAATCA AGTTAGCCCC AGCTTCTGCT TCAATGGTAT	2280
AGCCATCAAC AGAAACGTTA TTGAGCTGT CACACAAGAT GACAAATCCA CGAATCCCAC	2340
CATCACGAAC GATGATATTG CTTGCATTGC CAAGAACCAT CCAAGGGATA TTTTCTTGGT	2400
TGGCAAATTT CACAACGCGA GCCAACTCAA AACGATTTCTG TGGAAAGACC AAATAATCAG	2460
CCTCTCCACC TACTTTTGTA TAACTATAGC TATGCAAGGG TTCCTTAAAA CGGATATCAA	2520
TTCTTTCTAA GATTTCAAGC ATTTTCTCTC TTACAGACAT GTCACTCTTC CTTTACAAA	2580
ATTCAATCCA TTATACCATT TTTAGAGACA TTTGACGACC ATAAAAATAC CTTGTTTGGG	2640
TTTTGCATAA GAAAAAGAGG TTCCCCCTT TTTATGATTT TTTACAAAAG ATTTCTTGG	2700
TTCCATAGGC GACCAGAACG AGCTCCAGTG CTAGAATCAC TTCAACCAAG ACTGGATTG	2760
TCAACCAGCC TACTTGGAAG AGAGATGGTG CCAGATCAAA GAAGGCATGC AAGCCATAGG	2820
CTGCTAGGAG ATAAATCCAT TTCTTCTGGC GAACAGCTTG GTAAACCCAA ACTGTCAAAA	2880
GTAATTGGAA ACCAAGCGCC AAGATTGCTT CAAAACCAAG CAAATAAATC TGCCAGACCG	2940
AAAGTGAAGT AATGGTTTTT AACATATTTT CAGACAGTAA TTGCAAAACL TGTGGATTCT	3000
GAGTTTGAAC TGCCGAAAGA ACAATGTAAA GATTGAGTAA ACTAGTAAGG CCTAGAAAAA	3060
TCAACTCCAA GCCACCATGC CCC	3083

(2) INFORMATION FOR SEQ ID NO: 139:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15363 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

931

CCGGAGGATA TTGACCACCA CCAAAAGCAG GGGGAAAATC GAAATCAACC AATAGTAGGC	60
TACTGCGACA CTGGTCAACT CACTATCTGA TGCTTGATAA TAATGCAAAA AAGCTTTTAA	120
TAAAGGTTTG TCTATCAGCT CTTTCCACCA CTTTTTCATG TCATACTCCT TCACTTATAA	180
TCTTATACTC AATGAAAATC AAAGAGCAAA CTAGAAAAGCT AGCCGCAAGC TGCTCAAAAC	240
ACTGTTTTGA GGTGTAGAT AAGACTGACG AAGTCGATCA CATACATACG GTAAGGCGAC	300
GCTGACGTGG TTTGAAGAGA TTTTCGAAGA GTATTAACCTA ATTTCTTCTT ACCAATTCCA	360
CCATATCATA CGGTAGGGTA TTGGCAGCTT CCTTCAAGGA ATAGTTCTCT AAGTTATTTA	420
CATTTTGTCTG TAATTTCTTG GCATACTTAG TCGTAATCAA TCGTTTTTCT TCGTATTCGA	480
AAATCAACTT GCGCTCCAGA TAATAGCCTC TCAGCATTTT ATCGATATTG TTGGGTTTGA	540
CACGATTGAT AACCCGTTTG ACAAAGGCAC CACTGCTGAT AATAGCTGTT TCTCGAAGAC	600
GAGACTCCTG CATAAACTA ATCAAAGAGC GTCTGTAGAC TCCCTTCAGG TTTTCCAAAC	660
TTTCAATAAT CATCTCTGTA TTGGCAAGAT AGAGCTCTGC AATTGGTCA TAATCAAGAG	720
CACGGAGACG GCTTTGCTCC TTGTTCTTCC AGCTACGGAA GGTCTTTCCG AGAGTAAAAA	780
CTTCATGAAG GAGAAAACGT AAAATCCTCA AGGAAACAAG AAAATAATAG GTCAGTCTTG	840
AGGCAAGTTT ACGATTGATT CCTTGTCTA TATTTTTCAG ATAACGTTGG TAAACTCGGT	900
AAGCACGATT GCTAATGTTT CCCTCTTCAT AGGCCTGTTT CAAACCATCA CTTTCAATAC	960
TAAGAATCAA GAGTTTCAAA GCAGCCCAGT CTTCTTGATC ATCCTGGTTT TCTTGGCTTA	1020
AAATGAGATT TTCAATACGT CCATGATAAT TGTCAATAGC CGCATAGAGG GGAAGTTTAT	1080
TTCTGGTGTC TTCCAACCTT TTTTCCAACCT CTAGCGTTAC TTCATTCAAA ATGGCGATAT	1140
GCATAAGATA ATCCTTGCTT TCTTCTCTT CATCAGAAAG ATGAGGCAAG ACCAAGAGAC	1200
CTGTAAAAA GCTAACAAGC GTCACACCTG CAACAAGGAA AAGCAAAAGA GGATACTCCT	1260
GTTCTAGATT ACTTGGTATC AAGAGAATCG TAGCAATCGA CACCGTTCCC TTAACACCTG	1320
AAAAGGTCAA GAGAAACATG TCCTTCATAT ACTTATTTAG CTTTTTCTTG AGGCGTCGGG	1380
TTCTATAGGC ATAATAGCCA TAGATCATAA TAAACGAAAT GACAAAAAGG ACAAAGGTAA	1440
GGGCGATAAG AGATAGCAAT AAAAGTAGAG GATTATAGAT TGGATTGGTC AAGATAGGTT	1500
CTGCTATCAT TTCCAACCTC ATCCCTAAAA TCACAAAGAC AGAACCGTTG AGCATAAAGG	1560
TCACTGTATG CCAGACCGTC TCGGTCACCG TATCCACTTG GGCTTCGAGG AGCGTGATTT	1620
TCTTGAAGCG ACTTGCCTTT AAAATTCCAG CAACTACGAC GGCAATAATA CCTGAAACAT	1680
GAACTTCTTC TGCCAGAAAG AAGGTCACTA GAGGCAAACT CAATTCTAAT AAAAGTTCAC	1740

932

TGGCAATATC CGTTGCGCGC ACACTTAGCA AGAAGGTATG GAGGAAGCGG TTGGTCATGG	1800
CTGTTAAAAA TCCAATTAAA AAACCGCCTA GGATTGAAAA GATGAGCGAA CTGCTAGCTT	1860
GCCCCAGAGA AAAAGCTCCA GTTGTCCAAG CTGTCAAAGC TACCTGAAAA GCCACCAAAC	1920
CAGAAGCATC ATTCAAGAGT CCTTCGCCCT TAAGAATATT GGACACGCGC TTAGGAAAGC	1980
TAAAACGCTC CGAAAGAGAG GCAAAGGCCA CCAAGTCCGT AGGACCAAGG GCTGCCCCAA	2040
CAGCCAAGCA AGCTGCCAAG GGAAGGCTGA ACCAAAGAAG ATGGGCCAAG CCACCCAAAC	2100
TCAGGGTCGA GATAAAAATC ACTGGAATA TGAGATAAAC AATGATTCGC CAGTGTTTTA	2160
AAATAGCCGT AACATCTGCT TCTTCAGCCT CTCGGAAAAG CAAGGGTCCG ATAACCAGTG	2220
CCAAAAACAA CTCCTGATTA AGGTGAAAGT CAGTATTGGG TAAAAAGAGA CCAATCACAA	2280
TTCCCAAAAG AATTTGCACC AAAGGGAGAG GCAAAAAGGG CAGGAGCTTA TTGGTTGTAC	2340
TTGAGACAAT CAAAACCAAGT AAAAATAGGA TGAGGTAAAT CAGTAATTCC ACGCACGTCC	2400
TCCTTAATCT TTTTACAAC AGGATTCAAA TATCTCCTTC TGCTCTTGA TTTTGTGTC	2460
AATCTTGAA CAGTCTTGT GCTCAATTTT TCTCTGGCAC CGTTCCATT CAAGAGCAAC	2520
TAATTTTTTC TTGATTTAA GCATTTTTT GTCATATGC GCTTGGTCTA GCACGCCCAT	2580
CGCTCGTTCG TGGTGGGTTG ATTCAACAAA ATTCTGGCGC ATGGCATCCA GCTTTTCGTG	2640
TAAGTATTGT TTATCCATGT CTGTATCTCT CTAATTTTTT AATCATCACT AAAACGGCG	2700
GCTTGTGAC TTGGTTTAAA GTTCGGTAAA TGGCAGCTGT GTACTCTTGT TGTTTCAACT	2760
GGATCACAAA ATCCAAGACA GCATCTCTCT CGAGATCGCC TCCTTCATGA CCATAGTAAA	2820
TCATAATAGC AATTCGTCCA CCTTTGACAA GTAAGCCACA TAGCTTTTCT AATGCCTCAA	2880
TCGTTGTCTG CGGTCGGGTG ATGACAGACT TATCAGCTGC CGGCAAAATAG CCCAGATTAA	2940
AAATCCCTGC CTTAGCTTTT ATCACAAACT GGTCCAGTGT CTCATGGCCT TGCAAGATTA	3000
ACTGGGCATT TGTCAAGTCA GCCTGATGCA AACGCTCTTG GGTCTTTTCC AAGGCTTGCT	3060
TCTGAATATC AAAGGCATAG ACTTGCTTGG CTAGCTTGGC TAAAAAAGC GTGTCATGAC	3120
CATTTCCCAT AGTCGCATCC ACTACGACAT CCTCTTTTGT CACGACCTCA GCCAAAAAAT	3180
CATGTGCCAT CTCAAGTGGT CTTTTCATTT TCAAACCTCT GTTTTACAGC CTTGCATCCT	3240
TGAACACTTC CACGACGTCG CATCTCCATC TCAATGCTGT TGAGGACTTC CCATTTATTG	3300
AGGCTCCACA TAGGACCAAG CAGCATATCC CTAGGCGCAT CTCCTGTAAT TCGATGGATG	3360
ACGATATGTT TGGGAATAAT TTCCAGTTGG TCACAGATGA CCTGACATA TTCGTCCTGA	3420
CTCATCAATT GTAAACGCCC CTCATGGTAA TCTCGTTGCA TACGAGTATT TGTCATAAGA	3480
TGGAGCAAAT GCAGTTTAAT CCCTTGAATA TCGTTATCCG TGACACAACG GCGGACATTT	3540

933

TCAACCATCA TCTCATGGGT TTCACCAGGC AAACCATTGA TCAAATGGGA AACAACTCTCA	3600
ATTTTGGAT ACTTTCTCAA ACGCTTGACC GTTCCACCT ACAATTCATA AGAATGCGCA	3660
CGGTTAATCA GGTGAGAGGT TGCTTCATAA GTAGTTTGCA AGCCCAATTC AACCGTCACA	3720
TGCATGCACT CCGATAACTC AGCCAAATAT TCGATGGTTT CGTCTGGTAA ACAGTCTGGG	3780
CGCGTTCCAA TATTGATTCC TACCACACCT GGCTCATTGA TAGCCTGTTC ATAACGCTCT	3840
CGAATAACTT CCACCTTTTC ATGGGTGTTG GTAAAATTTT GAAAATAAAC CAGATACTTC	3900
CGAACATCCG GCCACTTGCG GTGCATAAAG TCAATTTCTT TATAAAATG CTCACGGATA	3960
GGCGCATCCG GTGCCACAAT GGCATCTCCA GAACCAGAAA CCGTACAAAA AGTACAGCCC	4020
CCATGAGCCA CAGTCCCATC ACGATTGGGA CAATCAAATC CCGCATCAAT AGGGACTTTA	4080
AAAGTCTTTT CTCCAAAGAG TTTTCGATAA TAATCATTCA AGGTATTATA AGATTTCATG	4140
ACTTTCATTA TAACAAAAAT CACCCACAAT CTCAAAAGCC TGACTTTCCT ATAAATTCCT	4200
CTGTTTCTCG TTTCCATTAG CCTTTTTTTA TGATACAATA TGGGTATGAT TTTAATGAAA	4260
TTAGCATCTA TTTTATTATT GATACTGACC TTAGTCGTCT GCATTATCCT AACCAAACCT	4320
TTTAGATTAA AAAAAGTAGG ACGAACTTT GCGGATTGG CTTTCCAGT CTTGGTATTT	4380
GAGTATTACT TGATTACAGC TAAACCTTT ACCCATAATT TCCTCCCTAG ACTGGGGCTA	4440
GCCCTCTCGA TCCTAGCCAT TATTCTCGTC TTTTCTTCC TTTTGAAAA ACCGAGCTTT	4500
TACTACCCTA AATTTATCAA ATTCTTCTGG CGTGCAGGAT TCTTATTAAC CCTTATCATG	4560
TATATAGAAA TGATTGTTGA ATTGTTCTTA ATGAAATAGT CGAATCCCTA AGCATTCTCT	4620
AGGGATTTTT GCTTCTCTA CAAAATAGTA TAGACAATAA CACTATACAA TTTTATACAA	4680
AGAAAAGAGT CTGGGACAAT AGTCTCTTAT ATCCAAAAAG GCAACGGATT TGCCGTTGCT	4740
TTTTTGGATG GTTACGATAG TCTTGGTAAA ATAGAATTGC CCAATAAACC ATTTAGAAA	4800
GCTATCCCAT GCATATTCAC TATAACACAA ATCAAACAAC TTTACCTA GAAATCAGTT	4860
CCTTCTTACC ACAAGATCAT CTCGTTTTTA CTATTGAAAA AGTGGTGAAT ACCTTGGAGG	4920
AACGTCACCT CTACACCTCC TATCATGCCT TTGATCGCCC GTCTTATCAC CCTAAAATGC	4980
TTGTATCTAC TCTTCTATTT GCCTATTAC AAGGGATTTT CTCTGGTCGA AAAATTGAAA	5040
AATGGAAGAG TTAGTGACCT TAGATTGTTT GTTTATTGAC AGAACTAAGA TTGAAGCCAA	5100
TGCCAACAAAG TATAGTTTTG TGTGGAAGAA AACGACAGAG AAATTCTCCG CCAAACCTCA	5160
AGAACAGATA CAGGTCTATT TTCAAGAAGA AATCACTCCC CTTCTGATTA AATATGCCAT	5220
GTTTGATAAG AAACAAAAGA GAGGGTATAA AGAGTCAGCT AAAAAGTTAG CGAATTGGCA	5280

934

CTATAATGAC AAGGAGGATA GCTACACACA TCCTGATGGC TGGTATTATC GTTTTCACCA	5340
TACCAAATAT CAGAAAACAC AGACAGACTT TCAACAAGAA ATCAAGGTTT ACTACGCCGA	5400
CGAACCTGAA TCAGCCCCCTC AAAAGGGACT GTATATGAAC GAACGCTATC AAAACTTGAA	5460
AGCTAAAGAA TGTCAGGCGC TTTTATCTCC CCAAGGTAGA CAGATTTTCG CTCACGCAA	5520
GATTGATGTG GAACCTGTCT TTGGGCAGAT AAAGGCTTCT TTGGGTACA AGAGATGTAA	5580
TCTGAGAGGG AAGCGTCAAG TGAGAATTGA CATGGGATTG GTACTTATGG CCAATAACCT	5640
CCTAAAATAT AGTAAAATGA AATAAGAACA GGACAAATCG ATAAGGACAA TCAAATCGAT	5700
TTCTAACAAAT GTTTTAGAAG TAAAAGTGA CTATTCTAGT TTCAATCTAC TATACAATAA	5760
GAGAATGACT CAAAATTAAA AAGCTAGACT TCCACAATTG GAAATATCTA GCTTTTGT	5820
GGTTGAGAAC TATTTTGTCT CAGGCTCTTT ATCTTCTATT TAGGACAAGA GTTTTCTTT	5880
GGTCTTTAAT GATAAAGAAG GTATCAAAAT TTCTAGTCTT CTTTTTTACC TTTAGTAACT	5940
ACTAATCCTG CACTCAAACC TAGAAGAGTT AAACCTGCTG CTACTGCTGC TTGGCTTGC	6000
GCACTACCTG TACTTGGTAA CTGGGCTTTA TTAGTTTGAC TAGCTTCACT TGAATCAATT	6060
GGTTTTGTAT CTGCTTTTTC TGACACTTGT GGTTTTTTAG CTTCTTGAGC TACTGGTTTG	6120
GTTCACCA AGACGATGCG GTCTGTCGGA ACTTCTACCA CTTACGGAG TTTTCTTCC	6180
TTACTTCCAT CAGGATTAAT CGCTGTAAAG ATACGTTCTT TTCCAACTTT TCCTTCTTGT	6240
TCTACACGAG TTTCACCTAG ATACAGTGT GAATCTTTT TCTCAACTGT CTTGTATGCC	6300
AAATCTTTT CAACAAATTC GATTTTGGGA AGATCTTCTT GTACAGCAGC AACTGTCTTC	6360
TCAGAACTG GTTTTCTCTT AGTCAAGTGG ATACGGTATT CTTGACTTG TTTTCCACTT	6420
TCTGAAACGA GGCGAACAAG TACTGGAAAG CTATCTTCTC CACTATCTAC CACAGTTGAA	6480
GCTACTTGAT TGTTCCTTC AACTGAGACT TTTGGCCGT GACCTTTATA GGTAATTGA	6540
TAGTCTTGAC GATTTTCAGC GAAATCAGCA AGTTCTTTTC CATCTACAAG AATCTTTGAT	6600
TGAGTGCTTT CTTGAGGCAA TTCACTTGGT GCAAGGAAGG TCATCTCAAT CATCGCAACA	6660
CCGCTCTTAT CTGCTTTACG CTCCATACGC CATCTCATAG CTTTGGCTTT GATAGCTTTA	6720
AATGTTACGT TGATTTATC ACCAGCTGCA ATGTCTTTAT CCGCACGATA AGGAACAGCT	6780
TCCCAATTTT CTGGATTGTT GAATGGATGG TCTCGTCGT AGGCTTGGTA GTTTGAATAG	6840
TAGGTTGGCA CTTCAAACCTC TGGACCGACA TAGCGTTCTA AAACGAGTTT AGATGGTGCA	6900
TCCGTACCAC TATCTGCAAA GAACTGAACT TTTCTTGTG TAACAGTCCG TTCTACAATC	6960
TTACCATTTT CACGGAAAAT CACACCCGCT GATACTTCTG GATTAGAAGA TGGTGTGGT	7020
GACCAGTTTG TCCAACGACG ATTTTCTGAA TGATCTCCGT CATTGAGATA GTCAACGCGG	7080

935

TCATGAGAGT TTTTGTCAAT ATCATTGGTT GCTGAAGCAA AGGCCTGGTT ACTGTTTTCA	7140
TCATAGTTAG GGTATCTGA AAGAGTCTCA CCAAGTTTGT CTGTCACTCG TACAGTGATC	7200
TCAGCAACAA GGTACTACC AAGGACACGG CCTCGAACAG TAAATTGACC TGCTTTTGTC	7260
AGATTTTCCG CTGGAAC TTC CATTCA ACTGTCAGGT CTTTGTTC GTAGCCGTCT	7320
TTACCTGTGA AGTAACTGG AACCTTAGTC GGCAATTCAA GTGCTTGACC TACTGTAGC	7380
AAGCGAGCTT GTTTAACCGC AGCAACTGGT TTATGAGAAA GTAAGCTCTT ATCCTTAGTG	7440
AAGTGCAGAC GGTATTCTCC TAAGATGTCG CCATTTTCAG CTTTCGGAT GACACGAACT	7500
GGCTCACCTT CACGAACGCT TGAACGACG GTAGCGAGAC CATTGTTGCT AACACTTGCT	7560
GTGACTGCCG GAACTTTTCC ATCTACAGAC TCAAGGTAGT AGTCTGTCAA ATCAGGGTTG	7620
AAGTTTGCTA AGTCTTTGCC GTCAACTGG ATTCTGTTT GTCCTTGCTT GGCTGCCGCA	7680
ACTGTTTCG CAAAGATTG TACCTCTGTG ATAGACGTT CACGCTTGT ATCTGCTTA	7740
ACCATGCGAA TACGAACAGC ATAGGTTTCA ACTTTATCAA AGCTAAAGTG GTTCATTTCT	7800
CCAGCCTTGA GTTGAGCAGG GGCTTTTAGA TTAGTAACTG GTTCCAGTT GGCAGAATCA	7860
TTAAAGACAT GGTCTCATT ACCAACAAAA CTAGGGTTTT TAGGAGCTGT TGGGACAGTC	7920
TTACCAACAT AATACTCAAT CACATAAGAC TTCGGTACAC CAACTCCATG GTCTTCATGG	7980
AATCCGACAC TTAGATTATC AACGGAGCGT TTGCTCAAGA TACCTGAATC TCCAAACAGA	8040
ACACCGACTG AAGCTTCTGG ATTAGTACGA TTCCAGTTG TCCAACGATT GGCTGGTTGG	8100
TTATTGTAGG AAATGAGCTT GTCATTAAACA TTTGAAACTG GGTGCTTGG ATTTGAGTCT	8160
GAAGCAAAGG CAAGTGCCAA TTCTGAACCG GTCCATTGGT CAGAAATGTT TGCACCTTGC	8220
TCAGTTTGAG CAGATACGCG AACATGAAGT TTAGTTGTTA ATTGCGTACC TTCTAAGCGA	8280
CCATTAACTG TAAAGACACC TTCCTTAGCG TATTGCTCTG GACGAATCGC ATCCCATGCA	8340
ACCTTAGCTG ATGAAACGTG ACCATTTGAA TCATATGTCC GAACACTTTC TGCTAATTGT	8400
GGTGCTTCTG CGATTGGAGT TGTCACTG ACTTCTTCAA CTGAAACGAT ACCTTCTACA	8460
GAGACTTTTG CACGCGCTTC AAGGTCAATT CTTCAACTT TACCTAGTAC TTCAAATGTT	8520
TGATAGGAGT CTAGTTTTTC TTTCGGAATA GCTTGCCAAG TGACTTTATG AGTTTTAGGG	8580
AAACCTTTGT CATACTCAAC TGTTACTGTT GCTGGAAGAC TTGGTTCTG ATGCAAATCT	8640
GTCACTACAT TTACAGGACG GATGGATTGC GCAATCTTCT TCTCAGTATT GGCTTGATA	8700
GTGAGTTCAA CTTGGTCTTT AGCTCCCTCA TATTCAGCGT TCAGAGTGAC TGCTCTGGC	8760
TTATGCAACT CAAGCATTC TTTACGAATT GCGACTTCCC CTTCACTACT TGTAGAGAAG	8820

936

GTTACTTTAT CAGCTGGTAA TACAGCTTGC GTTCCATCTT GATAGTGAGC TCGAACCGAC	8880
AATTTGACAG TTTGGTCTTC TTTGAGACTG TCAGCTTTTT CCACTTGCAA GCTCAAGTGA	8940
GCAATTTTGT GCGCTTCTTC AAGGAATTGA ATTGCATAGG TTTGAAGAGG GCCACCATCT	9000
TTAGGCTGAA TAAAGATGCT CGCACGCATG CCGTTTGCTG CGCTTGCTTG AAGAACTGTA	9060
ACAGCTGCAT TTTTAGCACT TGCTGTGACT TCTGGCAACT TAGCTCCATA AGCAAGAGTG	9120
CGGTATTGCA TTGGTTTMTG ACTAGTAAGA CCGTTTACTG CCTCACCACC AACCCTTACA	9180
GTTGGTACTG CAGGTGCCGC AGGATTGCCT TCTTCTACCA CAAGGGTTGC ATGAATTGGT	9240
TGACCTTCTA AATAACCGGT CGCTTGAATA CGAGAACCTG GAATTGCTAA CTTAGCTTTA	9300
TCTTCTTCGG CAATCTCCCA CTTGTCCACT TCATACTCTT CAACACTTCC ATCAATCAAA	9360
ACATAGGAAA CAGATTGTG TACAGAAATC AAGTCAGTAT TTGGAGCAAT ACGTTTCACA	9420
ACTGGTAGCT CTGATTAAAG AGCAATCACT TCTACACGAG CTTCTACTTC TCGTCCGTCA	9480
GCCATACCTT TCACCCTTAC AATACCAGGC TTGCTCACAT CTAAGAAGA CCAGGTTACA	9540
GGACGTTCTG CACGGCTACC ATCACTGTAT ACAAACGGAA CAGTGGTAGG CATTTTCAGGT	9600
GCCTCTCCAA TAATGGTCTG TACTTTTGGC ACTTCTGTCC CCAAACAGT CTTCTCTTGT	9660
CCTTCTTTCT TACCAGTAAA GACAGTGAAT TGGTTCGATT TCAAGAGATC AGAGTGGGCA	9720
GTCAGGGTGA ATTTCCCTGC TTGTTTCAAT GATTGACAA TGGCAACACC TTTACCATTA	9780
AATGCTTTAC GAATCCAAGA ACCATCTGCT TCGGCCTTAT AGCGTTCACG GCTGGCTTGT	9840
TCTCCGTTAT CTACACCGAC CAGTTGACCT TGGCCATGCA ATTGGAAGCG AACCAGATTA	9900
TTAGCAGTTG GAACCACATT CCCCTGGCTG TCAACAATTT CATAGTAGAT GTAAGTCAAG	9960
TCTTTTCCAT CTGCTGCAAT CGCATGGTCT TCCTTAATAA GACGAACTGC CGCTGGCTTA	10020
CCAGCAGTCG TAATCTTATC TCGAGCAATT TCCTTGCCAG ATTCATCAG AGCAATTGCT	10080
TCCAAGGTAC CTGTTTGATA GGCAACTTTC CATTCAAGAT AAAGTTCATT AGCATTTGCA	10140
CCTTCTTGGT AAGTCCGCCC ATCGCTGGTT TGTTTTTTAT TGAAAGTCTT AAGACCAAGA	10200
GATTTTCCAT TCAAGAACAA TTCTACACTA GAAGCATTTC AATAAGCACG AACTGGAATC	10260
TTACCTTCTG AGTCAGCTAC TTTGGATGCT AATTCTTTGT TTTCCAGTT CCAGTGAGGA	10320
AGAAGGTGTA CCATCGGTTT CTTCTTAACA GAAACCCATT GGCTTTGGTA GAGATAGAAG	10380
TCATGTTTTG GAATGCCGGC TGTATCTACG ATACCAAAGT AAGAGCTCTT AACAGGAGTT	10440
TGATTTTGGT TGTGCCATGG TGTAGGTTCA CCAATATAGT CCGTACCTGT CCAGATAAAC	10500
TGTCCAGCAT AGCCAGCGTT GTCACGGTCA AAAGTCCATG AAGCGGTTGC TGTTTTCCCC	10560
CAACCCACAC GATCATTTCC ATAATCTGAC TGTTCATAAT TACGCTCAGG TCCATTGCTA	10620



TGTTTCAATT CACGTTCCAGG GCGATAGTAA CTTCCACGTG TACGGGTAGC TGAAGATGTT 10680  
TCTGATCCAT AAATCAACCA TTTTGGATGC TTAGCTCTAA GGGCTTTGTA ATTATCTTCA 10740  
GAATAGTTAA ATCCAACAGC ATCGAGTTCA TCAGCAATTT TCTCATGCCC TCCGCTACCA 10800  
TTACCGAAAC GGAATTTATC TGCTCCCATG GTAACATAGC GAGTCTTATC AACATCCTTG 10860  
ATAACCTTAA CCAAACGTTT AACAGTTGCT AAAGAGTGGG CATCACCATT AGCTTCACCT 10920  
-----  
ATTTCAATTAC CAATTGACCA CATGAAGATA GCAGGGTTGT TTTTGCCTCT TTCGACCATG 10980  
GTACGTAGGT CAAAATCAGA CCATTTTTC A CTTTTCGAG CTTCTGGGTG AGTGGCATCT 11040  
TTTTCAAAGA AACGTCCATA GTCATAAGGT TTCTTGCCAC CATACCACGT ATCAAAGGCC 11100  
TCTTCTGAA CGAGTAAACC TAGTTCTGCT GCGATTTGCA AGGTTTGCTC ACTAGCAGGG 11160  
TTGTGGGTTG TACGGATGGA GTTAACTCCC ATCTCCTTCA TTTGTTGAG ACGGCGATAT 11220  
TCTGCTTTAT AGTTTCTTC TGCTCCAAGC GCCCCATGGT CGTGGTGCAA GGATACTCCA 11280  
TGGAATTTAA TACGTTTACC ATTCAAAGAG AAACCTTCAT TTGGAGTCCA GTGATAGTAA 11340  
CGGTAACCAA ACAAATCCTT CTTAGCATCA ACCAATTGAC CGTCACGGTA AACACGCGTA 11400  
ATCAATTCTG ACAAGGCAGG TTTGTCATTT AAAACAGTCC AGAGTTTGG TCTTTCAACT 11460  
TCTAAAATCG CATCTAGGCT TGTGATTCA TGTGCTTTTA AGGTACGACT CGCTGTACGA 11520  
ACTAAGCCTG TTACAGCATG ACCACCTCGT TCAACGATTT GATATTCCGC TACAAGTTCA 11580  
TGGTCTTTGT CGTCCGTATT GACGATTTTG CTGGTCACAT GAGTTTCAAC CTTGCCATGT 11640  
TGTTGTTCTT CAAGTTTGG TGTTAAAATA GTTGTCCCAT TTTTCTCAAC ATGCACCTTA 11700  
TCTGTCACTT GTAAAGTCAC ATCAGGATAG ATACCACTTC CTGAATACCA ACGGCTACTT 11760  
GGCTGTTTGT TGACTGCATG GACAGCAATC ACATTCTCAC GACCATCTTT TTGAAGGTAT 11820  
TTGGTGATAT CATATGAGAA CTGGTTATAA CCATTTGGAT AATGCCCCAC TAACTGACCA 11880  
TTGACATAAA CTTGAGAATC CATGTAGACG CCATCAAAAG TAAGGCGAAC ATTTTCTTG 11940  
AGGTCTTTTT CATCTAGTTT GAAAGTCTTG CGATACCAAG CTTCCCCACC GTTGAGCTGT 12000  
CCACCTTCAT TTTGTGCAGG AGATTCATGA TCGAAATCGT TAAAGATACT CCAGTCATAC 12060  
GGTAAATCTA ATTTTTTCCA CGTAGATACG TCTGCATCAG GTTTAATGGC TTCCTTAGAA 12120  
TTTGCAATGA GTTTAAAGTA CCAATTTTGA TTAAATCCA CTTTCTGTC TTCAATCATT 12180  
TGATTCACCT CTTCAATTGT TACAGCTTTA GCATCTTCCT TGAGCGGTTT TTCTTGATTT 12240  
GAAGCTTGTG ATTCTATCCT TGGAGCTTTT TCTTCCGGTT TAGCAGACAC TTTTCTCT 12300  
TTTGGAGTTA CGGCTTCATC TTCTTCTTC TCAGATGCAA TAGCCTCAGT TGAAGTAGGT 12360

938

TCACCTTGTT CTGTCCTTC AACTATATTT TTAGTTTCCA AAGCTTTATC AGCCTTTTCT	12420
TCTACTATCA TTTTTTCCTC TTTAGGTTTC TCAGCAGTAT GAGTAATAAG TGTTCATCC	12480
GCATAAACTA CAGATTCTCC AGCTATATTT CCTCCTAATA AACTGCACA AGTCCCAATC	12540
ATTACTGAGC AAGCTCCAC AGCAAACCTA CGAATGCTAT AACTCTTTT CCGATTCCAA	12600
TGGCCTTTCC CCATAAAACC CTCCTTATAT TATATTTAGT GCAGTTAGCT ACTACCAAAG	12660
CCCAAGTGGT ATACATGGTA TGACAACCTA GTTTCACAA TTTACACTCT GCGAAAATCC	12720
AATTCAAAC TCGTCAGTGT CGCCTTGCCG TAGATATGAT TACTGACTTC GTCAGTTTCA	12780
TCTACAACCT CAAAACCATG TTTTGAGCTG ACTTCGTCAG TTTTCTCTAC AACCTCAAAA	12840
CCATGTTTTG AGCTGACTTC GTCAGTTTCA TCTACAACCT CAAAACCATG TTTTGAGCTG	12900
ACTTCGTCAG TCTTATCTAC AACCTCAAAA CTGTGTTTTG AGCAACCTGC GGCTAGCTTC	12960
CTAGTTTGCT CTTTGATTTT CATTGAGTTT ATATTTTATA GGAGCGCATT ATTTTGCTTT	13020
TGCTGCGTAC TCTTCGTTAC GTTTGATCAT TTGTTTCTG TACCAAGCAA AGATACCGAT	13080
ATAGAATACA AGGAAGACTA CTGCACCAAG GATTGCTTTG ATATCACCAG TTGTAGTGT	13140
ACCAATTGTC CAACCAAGAA GTTTTTCGAT TGGTCCTTCA AGAGTAGAGT GAGTAATCAA	13200
TTGAGTTTGG CTCACACCTT CTGGAAGGC ACCTACACCT TTAGCAAGTT CTGTTGCAA	13260
TGGTGCAATA AGTGACCTG AAAGAAGGAA GAGTGGCAAC AAGAGTTTC CGAAGATAAT	13320
CATACGGAGC AATTACCAC GAGTTACAAC CAAGAGAGCT GGAGTAACAC CCATAGCGAT	13380
GATACCTGCA AGTGGCAAGA TACCATTTC AACTTTTGAA AGAAGCACTG CTTCAATCAA	13440
CATGATTGGT GCAAGTACGT TGGCACAAGC CCAGATTTC GCACGACCAG CGATGAATGG	13500
CCAGTCAAGA CCGATATTGA ATTTACGTCC TTGAAGACGT TTAGTAGCAA CTTTGTAA	13560
ACCTTGATGAT AGTGGTTCTA CGGCTGCGAT GAACCATGAA CCGATAAGTG AGAAGAGTTC	13620
CAAAGATACA CCGGCAGTCA AACCAAGAGA CAACCATCCT TTGATAACAA GACGCCATTT	13680
ATCTGCATCT GCAACACCTG CAATTGGATG TGGAGTTCCC ATAATACCGA TAACGATACC	13740
AAGGATGAAA CCGATGAAGA ATTTAGATCC CCAGAAACCG ATTTTCTTGT TCAATTTAGC	13800
AGCATCAAAG TCATATTTAT CAAGGCCTGG GAAGAATTTT TCAAAAATCT TATCCAAAAC	13860
CATGATAACT GGGTTCATCA TGTAGTTCAT GTGAGTTGAT GTCATTGGTG ATGAACCTGG	13920
GGCGTTAAGA AGGTCATCAA ATGTAGGTTT CATCAAGTCA GAGTTGATAA TTTTCAACAC	13980
ACCGACAAGG ACGATAGCTG CTGTAGCAAT AAAGAGTGAA ACCCCTTGAC TCACACCATT	14040
GTTATCAGCA TACCATTAA TCAAGAGACC TGTGATAGAC AAGTGCCAGA TATCAAAGAT	14100
ATCGACATCA AGTGTATCTG TTTTCTTCAT AGCTAGCATC ACTATGTTGA CAATCAACAT	14160

939

GATGAGCAAG AAGTATAGTG TCCAAGCAGA ACCCCAAGTG ATTGTAGCAA GTGGTGCCCA	14220
ACCAACGTCG GTAATACTCA ATTGGATACC AGTGTMTTCA ACGAATTTTG CTAGTGATGC	14280
TGAGAAAGCA GTGTTTAGCA TACCGATGAT AGCACCGATA CCTGTAAGAG CGATGGCAAG	14340
TTTGATACCA CCTTCAAGCG CTTTGGAGAA TTTCACTCCA AAAAGTAAAG CCAATACTGT	14400
CAAAATGATT AACATGATGA CAGGTCCACC CATTTCTAAG ATGGGATTGA AAACCTTTCC	14460
GATTAGGTCA AAGATTGCAT CCATAACAGT TCCTCCCTTT TTGATGTTAT ATGAATGTTA	14520
ACAAATTAGA ATTAGCTTAA TCCGTGTTCT TTAATAGCTG CTTCAATATT GTCAAATACT	14580
GGAGCGCTCA TTGCTGGGAT ACGGAATAAG ATTGGCCCAG CTTGATAAC TGGGATACCT	14640
GGTTCAAAAC CAAGGTCTGT TGCAGCGATT GGTGTAAAGA TATCGTAACC TTTCATAAGG	14700
TCTTCGTTTA CATCTTTCAC CATGACTGCA TCACAGTGAA CATCATAACC ACGGTTTGAA	14760
AGTTCCTCTT CTAGAGCACT TTTAATTTGG TGACTTGAGT TAACACCTGC ACCGCAGGCA	14820
GCAAGAATTT TAATCATTTA GATTTCCTCC GATTTTATTT TTTAATAGAC AAGATTAAGC	14880
GGTTGCTTCA GCAATGTAAG TATAAAGGGC TTCTGGTTCA GAAATTTTTG ATAGGTCTTC	14940
AAGATGACCA TTTCTGTGA AGAAGTCCAT TAACTGAGCA AGAATGTTCCG TTTGACTTGA	15000
ACTTGAATTA TTAATGATAA AGAAGAGTAG GGATACTTCT ACTTCCTTAT CAGGAGCTAT	15060
CATATTGTGA AAAGTTATTG GTTTTTCTAA TCGAACAACC ACCACTTTCT CAGCTAGATT	15120
ATGAACAATA TCTGTGTGAG GAATCGCTAC ATTTGGCAAG TCCTTTCCTA GAAATTCAT	15180
ATCTAAACCA GTTGGAATG ACTTTTCACG CGTGATCAAG GCTTCACGAT AAGTTGGAGT	15240
GACAATTTCT CGTTCTTCCA ATAAAGTTGC AACCTGATCA AAGAGTTGTT CTTGACTATC	15300
CGCTTCTAAG CAAAACACAA GGTTTTTGTC AAAGAAATAA TCTAATACCA TAAGTTTTTC	15360
CGG	15363

## (2) INFORMATION FOR SEQ ID NO: 140:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 28882 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

TAAGACTATT TAATAGTGGG GTGAAATAGG ATACGAACAA ATTGATTAGG AAAATCAAAT	60
GAATTTATAG AAATCTTTTA GCAGTTATGT TATCCTATTC TAGTTTCAAA ACGCTATAGA	120

940

AGCAGCATTG TGCTAGTCKA GATTCAGTTT ACTATACTAA AACGAGTAGC TTGAAATCAA	180
AAAACCCACC CTCACAGGCA GGTTTTATCT GTATTATTCA GCTAGATTAT GCTTTACCTT	240
CTGAACCGAA TACGTCGATA CGTTCTTCAA CCGATGCTTG GATAGCTTTT ACACCGTCAG	300
CCAAGAATTT ACGTGGGTCG AAGAGTTTTT TCTTGTCGTA TTCTGCTTCG TTTGCTTCGT	360
AGTCACGAGC AAATTTACGA GTTGCCTTAG CGAATGCGAT TTGGCATTCT GTGTAAACGT	420
TAACTTTGGC AACACCAAGT TTGATAGCTG CTTGGATTMG CTCATCAGGA ATACCTGATC	480
CACCGTGCAA TACGATTGGG AATCCTGGAA GAGCTTCTGT CAATTTTTGC AAGTGGTCAA	540
GGTCAAGACC TTCCCAGTTT ACTGGGTAAG GACCGTGGAT GTTACCGATA CCAGCTGCCA	600
AGAAGTCGAT ACCAGTTTCA ACCATTGCTT TAGCGTCTTC GATTGGAGCC AATTCACCTT	660
TACCGATGAT TCCATCTTCT TCACCACCGA TAGTACCAAC TTCAGCTTCT ACTGAGATAC	720
CTTTAGCGTG TGCTTTTTCA ACAACTTCTT TAGCCAATTT AAGGTTTTCT TCAACTGGAA	780
GGTGTGAACC GTCAAACATG ATTGAAGTAT AACCAACTTC GATACACTCA AGTGCATCTT	840
CGTAGTGACC GTGGTCAAGG TGGATAGCTA CTGGTACAGT GATACCCATT GATTCAACAA	900
GGTTAGCGAT CAAGTTGCGA GCAACTTTGT AACCACCCAT GTATTTAGCA GCACCCATTG	960
AAGTTTGGAT CAAAACTGGA GCTTTTTTAG CTTCTGCTGC GCGCAAGATA GCTTGAGTCC	1020
ACTCAAGGTT GTTTGTGTTA AATCCACCAA CTGCATAACC GTTGTACGG GCTGCTTGGA	1080
CAATTTTTTC TGCTGAAACG ATTGCCATTT TATCAGGCCT CCTGTATATT TTTATGGGTC	1140
ATCCCATTTA CATGTTCAT TTTATCACTT TTTGCCAAAA AAATCTAGTT TTTCCCGCAG	1200
TTTCGATTGA TTTTCTTCTA ACTCCATCTA TGTAACCCT TTCTCTCCCT AGTCTTGGAC	1260
GACTTTTGGA AAATCTATAA AGAAGGTAA ACTATTCTCC TCCATCTCGA AACGATAAGC	1320
TAATTTTTCA TGTCTAATA GACTCTTAAC CACAAAGAGC CCCATACCAG ACCCCTTGAC	1380
CTTGCGACTG GCATTGTCAG AAAAAGACTG GGCTAGTTTT TCTTGTTCTT CTGAGCTACA	1440
GCTATTTTCG ATAAAAAGTT CTCCTTCTCT TTCTCCAATT CGAACTAAGC CACCTGGAAC	1500
AGAGTGCTTA ATGGCATTGC TGATGAGATT AGAAAGAATC AACTTCATAA CTGATGGGTT	1560
TAGATAAGCC TGCTGATGGG TCAAACTATT GTCTATCTGG AGCTCTCTTT CCTTGGCTAG	1620
CAAGGCATAA TCTTTGACCA GATTTTGCCT CATCTGGAGG AGGTCAATTG TTTCCCTATC	1680
ATCTCGCAAT TCCTGCACAG AAGAGAGGGA AAGTATCTGC AGAACATGGT GATTGAGTTC	1740
ATCCACAATC CCCAAGGCAA CTCCCAGATA CTGGTCTCTA TCCTTATAAC GACCGATATT	1800
CTCTCTCATA TTTTCGATTA GGATTTTCAA ACTAGCCAGC GGTGTTTTCA ATTCATGAGA	1860
AGCTCCTCGT AGGAATTCGA CCTTCATCTT CTCCAGCTGG AGAATGGCTT CATTCCTTTC	1920

ATGCAAGTCC GCAATAACAG TCAAGAGATG CTGGTAGAGG CTATTGATTT GTTCCTTGAG	1980
ATTACCTATC TCATCCTTAG AATCCACGCG CAATCGCACT TGGGAATCCA GGTCCATCAT	2040
CCGACGGGTC ACCCGCTTGA TTTCCAAAAT CGGTGCAACA ATAGTCCGAG CGTAGATGTA	2100
GGCCACCAAA AGGGAAATCA GAAAGGAGGC CAGCAAGGTA TAGGGAAGAA ACTGGAGACT	2160
GATTTGCTCC GCTTCCTTTT GTAAATCCAT GGAAGCTAGA AACTGGAGAA TCATAGTACC	2220
ACCGTCTTGC GTTTTCACCT CGCGCTCCTC AATAAGAGA GAGGTTGTCT GGCGGTCTGT	2280
GTCCAGAGGA AGACTGTCCT TGACTTCTAA CTGTCTCTCG GTCATCTCAC CTTTGACGGT	2340
CCCCTTGATA TCACTAGTCT GGGAAATACAA GTCTAACACT TGCTCGATAC TCTGCCTATC	2400
TTTCCCTTCT AGGGACTGGG CAATGGCTGT TGCCTTTTGA CCAATGGTTT CCTGACGATG	2460
ACTCAGATAA GTCGAAGGAA AAAGAAAATA AATAGCTAAA TGAAGGCAGA TAACCAGAAC	2520
ACTAAATATC GAGAAGGTAT AGATAAATAT CTTTGCAAAT AAACCTGTTT GTTTTCATTTT	2580
CGTCCCAATT TATAACCAAC ATTGCGCACA GTGAGGATAC AATCCAAGTC TAGCTTTTTC	2640
CGCAATTCCT TGATATAAAC ATCAATAACA CGGTCAAAGG GAACCTCATC TGTCGCTTTC	2700
CAGACGGCAT CGATAATCTG AGATCGAGTC AAGGCCCCGGC CTTTCATTTTT CACTAGATAG	2760
TCCAGAATTT CCAACTCTTT GGCATTGATA GGCACCTTCT GACCTGCGAG GCTTGCACTG	2820
TAGCTTTCAA AGTCCACCTT GGTATCCTTG TAAGAAAAGA TTCGTCCTGT ATCGTAGTAG	2880
CGCTTGAAAA TCGCGTCCAC CCTCACTTTT AAAAGGGAGA GGGAGAAAGG TTTTTCAGA	2940
TAGCCATCTG CCAAAGAGGC AAAGGCACTC ATCTTGTATT CCTCATCTTG AAAAGCTGTC	3000
AACATCAAGA CAGGAACCTG ACTGTTTTTA CGAATCTCAG CTAGGACTTC TAAGCCGTTG	3060
AGCTTGGGCA TCTGGATATC CAGTAAACC AGGGCCACCT CATAGCTAGA AAATTGCTCC	3120
AGAGCTTCCT GACCGTCCGC TGCCTCAATA GTTTCATAGC CACAATCCGT CAAATAATCA	3180
CTGACCCCTT CACGGATCAT CTCTTCATCT TCTACAATTA AAATTTTCAT ACTTTAACTG	3240
CTCTCTATTT TTTATTTTTC TTAGAATAAA TACCTACCCT ATTTTCTATT ATAGTCTCTT	3300
GCTGGCCTTT TGTCTGCAAG CAACTGACCA CTAGATAAAA CGTTGTGAAA TTCCTTTCTC	3360
ATAAATTCCA TAACTTTAGT ATATTATATT TAAGCACTAA AGTACAAAGA AAGCAACTGA	3420
AAGCAATGAT TTTCAACCACT GCTTTCGGAT TTATTTTGAA TTGTTAAATA GCCATTCTTA	3480
TCCACTATTC TTGAATAGAA ACACAAGATG CAATCTTTAT TCTAGACTCA TTTTTTCAAA	3540
TTTATTCACC ATCCAGCAAG AGCTCTTTTG GTTGTTTTCT AAGGAGATTG CTTGAAGCAA	3600
GCGCCATAAC GAGAACCACT AGAACCAAGG CAAGGACAAA AATGATGATA AAGTCTGATG	3660

942

TCTGAATGGA AATGTCTAGG CTCGACAAGG TCTTGCTAAA GCCATCTACT TCTGCACCAC	3720
CACCAAGGTT AGAGGCTTGA GCCGCCTTAC TAGCCTGTTT GGCAACACCT GAAGTCACAT	3780
TGGCAAGGAC AGTGTTCCTCA ATTGCACGGG CAGTGTAAAT AGCTAGGAAG TAAGCAGAAA	3840
CTAGAGCAGG GATAGCAATC AAGATAGATT CGGTGATGAA TTGACCCAAG ATACTTGCCT	3900
GCTTGAGGCC GATAGAGAGG AGAATTCCTCA CTTCCTTGCG ACGGGCGTTG ATCCAAAGGC	3960
TGAGCAAGAG GGCAAGGAGG AGAACTGAGA AGCTCAAGCT ACCCCAGAAG AGGAGGTTGG	4020
CCATCTTGTA CATACCAGAG ATAGATTGCT CAAGAGCTGG GTAGTTAGAG GAGCTCTTGA	4080
CGAGTGTGTA GCTCTTCAG TTGATACCAC TGATGCCATT CAACTCTTTC ATAACATCAT	4140
CCAAGTCTTT GTCTGCTGTT ACAAAGAAGG TTGCGTCCCC ATAAATGGCT GTGTCTTCTG	4200
TGTATCCATA AAGTTTTGCA GCAGTGTGAA TGTCTGTAAT AGCTGTGTTT TCGTAAAGTT	4260
CTTGAGTA GGTACTGCT GACTTATTAT GACCATCAAA GAGTCCCTTG ATTGTCACTT	4320
CAACTGTTTC CTTGGCTCCT TTTTCATTAT CTGCATCGTA GATATTAGAG TCCAGTTTAA	4380
CCTGTGCCCC TACTTTCAG CCGTGTGTTG CTGCCAAGTC CTTGTGCAAG AGGATTTTAT	4440
CCTGTGCGTC GTTGGTTAAG TGCTCTCCTT CGACTAGTTT ATAAGAACCA GAGACAACT	4500
TGCTTCTTT AGAGGAGTCA TTGACACCTG TAATCATCAA GCTACTTCCA AAACGCTTG	4560
CACGATCAGC AGTGAGATTC TTCTTGTTT CTGGCGTTTC AATCAGGTCA TATCCAGTCA	4620
AATCTCCGAT AGCGTTGATA CGTTTGACAT AAGACTCAAT GGCCTTGTTT TCGGTGATTT	4680
TTTGATGTC TTCACCCTTG ATATTCCCAG CACCACGAGG CGTTCCTTG TTGACCGAC	4740
GATTGATTTG CATGGAGAAG CTATTGGTGA TATTTTAAA GGTCTCCTGA GAAGCCTTG	4800
CAGTAGCTCC CTTGATTGAC AAGCCGACCA AACTCAAGCT CGCCATGAGG AGAATAATCA	4860
GGAAGATGAC AATCGATTG AAAAACTTCC TTGTAACATA GGCAAATGCG TTGTGTAACA	4920
TAGATTCCCT TTCTAGATTT TGTTTTAATC ATTCTATTAA AATAAGCTCA AATTATTTAC	4980
TAGTATTGCG CGTTTCAGTC AGTTTCTTAT CCTTTAATTC AAGTGTAATA TCTGACGCTT	5040
GTGCCACTTC TTTACTGTGA GTTACGACAA TCACACATTT ACCTGTTTTT TGGGCAAGTG	5100
ATTTGAGTAG TTCGACAATA TCTCCAGCAG TTTTAGGATC CAGATTTCCCT GTTGGCTCAT	5160
CAGCTAGAAT AACTGGAGCT TCTGAGACCA AACTGCGAGC AATGGCAACA CGTTGCTGTT	5220
GACCACCTGA TAACTGGAGA ACATTCCGCT TGATCTGGCT TTCATCCAAA CCAAGCTCAA	5280
GAAGTGATT CTTGCTTGCC TTTTGTGTA CCAATCGGAT ATTTTCCAGC GGAGAAAGAT	5340
AATCTATCAA GTTATAATTT TGAAAGACCA GGGAAATATG GTGCATGCCA TGGTAAGAAT	5400
AGCCCTTCTT ACGAATATCC TCTCCTTGAA AAAGGATAGA ACCTTCAACA GGACTATCTA	5460

943

GACCAGCAAG TAGGGACAAG AGTGTGGATT TTCCTGCTCC TGAATCCCCA ATAATACTGT	5520
AAAATTTTCC GGGTTCAAAA TTATAATTGA TCTGATATAG GACTGCTTCA GCAGTATTCT	5580
TATAACGGTA GGTAACATCT TGTAAATTGA ATAAAGTCAT GATTTCTCCT TCTTAACTAA	5640
TAGATGATAA AATTTCTTTC GGTGATTTTC TAAATAAGAA TAGGAAACAA AGGGCTACAG	5700
ATAAGCAACT AAGCAGAACT AGAAAAACAT AGGATTCTGC AAAAGATAAG ATGCTAGTTG	5760
ATAAACTGCT TGCTTTGGCT AGTGTATCTT GTAAGCTTGC CTGATCTCCA CTTGCTAGTA	5820
GAGTTTGGAG TAGGTAAGTT GTGATTGCGT TTCCTGCAAC AAATGCTGGA AGCAAAGCTC	5880
CAAGAGATAC CAAAACCTACC TCTAAACAGA ATTGTAGGAA GATCGAGCTC TTGCCTTTTC	5940
CAAGTGCAAG TAAAATCCCC ACTTCATAGA CCCGTTCTCT CAACCAGAGA GACAAAACCA	6000
GAATTAAGGC TCCAGCTCCT GCTATCAACA TCCCATAAAG GAAGATGGTC AGGAAGGTTT	6060
GGAAAGTGC AACTGAGTCT TTGATTGTT CAAAAGCCTT GTTTTCCTTT TCGACTTGGT	6120
AGCCTTGATT TTCCAAGGCC AAGTTTCTA CCTGCTTCAT GAGTCCGTCC ATTTCTTAG	6180
GATTTTCTAC ATAGAAGCGT GCTGCACTGA CTTGAGCTTC ACTATTGCCC AAAAGGGTTT	6240
GGCTACTTTC ATAGTCTGTA AAGACTTGAT TTTCACTGAA GTCAGAAGAC AAGCCTGTGA	6300
ATTTCTCTTG TTTTTTACCA GAAAAGATGC CGATAATCTC AACTCTACT GTTTGTCCTT	6360
TTCCAGATTC AGACTGACCA GCATCCAAGC CAATCTTGTC ATGAAGCGAA AGACCGTTCT	6420
TCTTAGCCAA TTCTTCGTGG ATAAGGATTT TCTTGAATC CCCTTTTGA AGGTGTCGCC	6480
CTTCTTTTAG ATTGAAAGCC GAAGTGTAA AGGTTACATC CTTGGATGAA TCCTCAAGAG	6540
CCGTAAAGCT AACCAAGTTA TTGTCTGCAG CTGATAAATC ATCACGCTCC ACGCTCTGCT	6600
CGCCAGTCAC TGCTTCCTTG TCTTTTAGTT TTGCGACCGT CTCAAGTTCA GGAGAGACAT	6660
TTCCAGCCC CTTAATCTTG CTTACAGATG CTAGGTCTGA CAACTTGAAT GTCTGACCAT	6720
TCTCTATCTT CTTAATAGAA AAAGATGTAT TGAGTGATTT ATAAAGATTG CTTTCTACTG	6780
TTTTGTTGGA CTTTCATCAGA GTCAAACAGG CTGAAATTCC GGCCAATAAG ACCAATAAAA	6840
TCAGAAATAA AATAAACTT CTCAGTCGCT TTCTGCTGAC ATAAGCCCAA GATCTTTGGA	6900
TTGGATTCAT TTGTCACCTC CATATTTGTA AGACTATTAT AAAACCCAAA TATGAAATAT	6960
TTATGAAATA CGAAAAAAA ATATCGAGTA GGGGATAATC TCTAGCCCTC CTCACACCAC	7020
CATACGTGCC GTTCGGCATA CGGCGGTTCA ACTAATTTT AACGCATGTC GTTCAAGGTA	7080
ATAATCCAAA CACGAAACCA GTCCACGTTT TTCAAGGACT GGTTTTGATA TAGCACGTTT	7140
AAGTACCGAC TTCTGAGCTA CTATAGTAGA TTGAACTAG AATAGTACAC CTCTACTTCT	7200

944

AAAATATTGT TAGAAATCGA TTTGACTGTC CTGAACAATT CGTCCTATTC TTATTTTCATT	7260
TTACTATAAT TGATACTGGT CGCCCCAGCC AGATACCTTA TCTGCTATCC ATTTAGGAAC	7320
CCCTAACTTA AGCAATCCCC ATAATCGTCT CGATTTCCTC TTCCATTGCT TCCAGATAAT	7380
CACTCGTAGG CGAGTACGCA AGCGCTCATC TATGCTAGTG ACTATACTTT TCATATTTAT	7440
AATTCATTCC TTTCTTTTCA CTCAAGGCAC AACACAGAAT GAAAAAGTGT TGTGATCTTT	7500
ATTTTGT TTTT ATAATAATAG TGAGAAAACC TATCACTACT ACAAATCACC GGGAGGTGAA	7560
TAAGTGAGTG GTACAGCCAC TACCTCGCAT ATTTTGT CAC ATCATTTAAC GGTACATAAT	7620
AAGTTGTACC ATCTGAATAA GTTGCTACAA TATCATTTGC ATGCTCTCCT TCACCTTTAG	7680
CAAAGGTTGG AGCTCCTGCT GGATGATTTT TATTTGCCTC TTTCAATTTT TCAATAATGG	7740
CATTTTTTCT GTATCTTTTA TATTATCAGG ATTTTTCCT AAGATTTTGT CTGGATATGT	7800
CGGTTTAGCA GAAACAATTT TTAAGTTTAC TTTCTTTTTA TTGGAAGCAC TTGTCCAGTT	7860
TCCAGCATT TCTTTAGCAT TTAATTTTAC AGTAATTCCT GAACTAGGAA CTTCAGTAGC	7920
AGGTTGATTA TCAACATTAT TCAACTTTAA TTTCAAAGA GCTGTTGCAT CAGACGTTTT	7980
ATCAATCGTT ATATATAATG ATGAATTGTT ATTATAAACA GTTCCTTCAT ATTTAGCTGT	8040
TTGTGAGCTA CTTGAAACAG AACTGAAATT ATACCCACTA CCTCCCTGAT TATCTTCAAT	8100
GCTTACGTCT AAATGAACCT CCCCCTATT ATTTGGCTTA GCAACAACCTG TTATAGTAAA	8160
ATAACATAAA ATTTGCATAA ATAGATTAGG GAAATCAAAG CAGCTTCTAG GAATGTTTTA	8220
GCAGTCACAG TGTACTTTCC CAGCATCAAG CCACTATAAC TCTGCACATA AAAATGGAGA	8280
AGATGGCAAT CCTCTTCTCC AAATATTAAC TTCTTTACAA ACCAACTATA GTTGACAAAG	8340
AACCTAAAAT CAATTGATAA CACAAGGTCA GGTGGGTCAA CTCTTTCAAC TGAAGCCCTG	8400
TCAACTCTTC CCATTATCA ATCTTGATT GGAGAGAATT GCGGTGCAGA TAGAGTTGCT	8460
GGGCTGTTTT AGTGAGAACA GCACTATTTT CCCAAAGAGA GAGAATGATT TCCTGAATCT	8520
GATCTTGATC CAAAATCATC TGGTGTAGAC ATTCCTTGAT TGGCTTCAAG TCCACGAGTC	8580
TTTCTCCCAT ACTCCAAAGA TAGAGCTGAG AAAAAGTATG AACACCTTGG TGACCCTGAC	8640
GCCACCATGT CTTGAACAAA TCCCGCTCAG CTTTGATTAA GTCTGATAGG GCTTGATGTC	8700
CCGTCTGAGA CCAAACCTGA CCCAACATGA TAGAAAGACG AAGTCCAAAG TCATACTCAA	8760
CCGCTTCAAT CGTATCACTT AAAATATCTC TTACAGAAGT GTATTTGTCT TGTGAAGCA	8820
CGAAAACATA ATCCTGAGCT CCGACCTGTA GCACTGTCTG ACAAATCGGA AAAAGAGTCC	8880
GCATCATATC TAGCCAAGAA GCCAGATTTT CCTGCTGAAA ATAAGAAAGA TGGCAATAAA	8940
CCAACTGAAT CTTTTTAAAA ACTTGCGGTG CCTGTCCCTT GCCCTCAACC AGATAGGAAT	9000



945

ACCAAGGGTT TAGCGAACGA GCCTGCTCCT GCTGGGTCAA AAGGGCAACC AACTGCTTTT	9060
CACGCTCGCT GAGCCCAGCT TCCTCCAGCA AAATCCACTG CTGAGAAGCT AAAGGGAGCG	9120
TGAGATAGCC CTCTTTCTCT ACTGGTTGGT CTGAAATCCG AGCCTCAGGA AACCAGTCTT	9180
GTAGTTCTTT TGCCCTCATG TTCTAGCCCT CCACTTTTTG GATGCACCAT GAAACCAAAC	9240
TCTCAAGACG TTCCAGATTC TCAGTCATAT GGAGATAGCC CATAACCGCT TCAAATCCCG	9300
<hr/>	
TGGACATACG ATAAGTCACG ACATCTGCAT TTTAGCCTT TGTGTGGCTA TTGGTATTGC	9360
GGCCACGTTT GTAGATTCTT TCTTCTTTTT CCGTTAGGAC CTGCTCCTCC AACATGAGAG	9420
CAATCAGGCG AGCCTGAGCC TTGGCTGACA CGTACTTAGT TGCTTCTTGA TGGAGTTTAT	9480
TGGGTTTGGT CATACCTTG AGGATGAGGT GACGGCGAAT ATACATAGAA TACACCGCAT	9540
CCCCCTCAAA GGCTAGCGCA ATCCCGTTAA TGAGATTGAC ATCAATCAGG TGTCACCTC	9600
ACTCCATCCT TGGTATCAAG GAGCTTAATT CCTTGAGTAA CCAATTGGTC ACGGATTTGG	9660
TCTGCTGTCG CAAAGTCACG ATTGGCACGC GCCTCTTGGC GTTTTTGAAT CAAGTCTTCA	9720
ATCTCTGCAT CCAAACTTC CTCAACAAAG ACAATTCCAA AAATTTCTAA CATATCTGCA	9780
AGAGCTTGCT TGACACTTGC ATCATAGTTC CCTGAGTTGA TCCATTTGGC CATTTCAAAG	9840
ACAACTGTGA TACCGTTGGC AGCATTAAAA TCTTCATCCA TAGCTGCTAC AAATTTATCT	9900
TTAAAGTTTT GTAACCTTG GGCATCCACA TTTCTGTAA ATGGTTGTTT GTAAGTATTC	9960
TTGAGATACT TGAGATTGGT CTCGGCATCG CGAACTGCCT TTTCCGTGAA GTTGATAGGC	10020
TTACGGTAGT GCTGGGTCGC AAAGAAGAAA CGAAGTACTT GCCCATCAAG AGTTTTAAGG	10080
GCATCGTGTA CCGTAATGAA GTTACCCAAG GACTTAGACA TTTTGACATT GTCGATATTG	10140
ACAAAGCCAT TGTGCATCCA GTAGTTAGCA AAAGCCTTGC CTGTTTTAGC TTCAGACTGG	10200
GCAATTTTCT TGGTGTGGTG TGGAACTCT AGGTCAGCTC CACCACCGTG GATATCAATG	10260
GTATCACCTA AAATCTCTGT CGACATGACT GAACACTCAA TATGCCAACC CGGACGTCCA	10320
GGTCCCCAAG GACTATCCCA AGAAATCTCA CCTGGTTTGG AAGATTTCCA TAGAGCAAAG	10380
TCTACAGGAT TTTCTTACG AGCCGTTTCT TCATCGGTAC GACCTGAAGC ACCTAGCTCC	10440
AAATCTTCCA AGGTTTTATT AGCCAATTTA GCATAGTTGT GGGATTTTTC TACACGAAA	10500
TAGACATCCC CTTGACTCTC ATAGGCAAAG CCTTCTCGA TCAAGTCTTC CACAAAACGG	10560
ATGATGTCTG CCATAAACTC CACTACACGC GGATGGCGAG TCGCAGGTTT CACGCCAAT	10620
GGCGTCACAT CCTCACGAAA GGCAGCGATG TACTTATCCG CAACCTCCTG AGGCGTGATA	10680
CCTTCTTCCC TGGCAGGTT GATAATCTTA TCATCCACAT CTGTAAAATT GGAAATATAG	10740

946

GCAACCTTAT	ACCCACGGTA	CTCAAAATAG	CGACGAATCG	TATCAAAAGC	TACCGTCGAA	10800
CGGGCGTTTC	CTACGTGGAT	ATAGTTGTAC	ACCGTTGGCC	CACAAACATA	CATCTTGATC	10860
TTGCCGTCCT	CAATCGGGAC	AAATTCTCGC	AAATCACGAG	ACATGGTGTC	ATAGATTTTA	10920
ATCATAAATC	ATAATCAGGA	AAGCTGAAAT	CCAAGAACAA	TTAGTTTCAT	CACTAAAAGT	10980
TCAAGTAAAT	TTCAGTCCGA	ATATCTCTAC	ACTTCGGAAT	CEETTGCTCC	TTTCTCATTC	11040
AGATAAACCA	CCTGAGTCTG	TTTGACAAAG	CCAATTTTTT	CATACAAACG	TTTGGCACCT	11100
ACATTGCTAT	CTTCCACTGC	AATCTGAAAT	TCCTTGTCAT	TTTGCTCAAT	TAGTTGGTTG	11160
ACGAGGGATT	TTGCTAAGTA	GCTTCCATAG	CCTTTTCCAC	GTTCAGGTTT	CAATATTGCT	11220
AAACCGTAGA	GGTAATTCGT	ATTAGTCGAT	AAATCAACCG	TACAAGTTCC	AATAACCTGA	11280
CCAGCTTTTA	ATAAAATATA	TAGTCGGCTT	TCTGGATCTT	TCAGAGCTTC	AGCGACATAT	11340
CTATCCACAA	CTTCTCTCGA	TTTATGTTCC	TCTGAAATG	CCTGAAATTT	TAATTGACTA	11400
ATTTGATCCT	GATACGAACT	ATCTGCTAAC	AAAACCTCAA	GATGGGAAAC	ATTTGCTAAC	11460
GGATAAGGTC	TTCTATCCTT	ACCTAACCAA	GTTTCTGTCT	CTTCATCCTC	GATTAGTCCC	11520
CAGTTACTGG	CAAAGTCAGG	ATGATTCTCT	AAAAAATAC	GTTCTGTCTG	AAAAGTGAAT	11580
GACCGAATGG	GGAAAGAAGC	TGTTTCTCTC	TCAAACTAG	TAAACAATGC	ACGCGCAATC	11640
CCCTGACGGC	GATGACCTGG	ATGAACCACT	ATCGTCACTT	CTACATCTTG	GTCATCTGCA	11700
TAGACGTTA	ATAAACCAAC	AAGTTGCGCT	TTTTCATAAT	AAAGGAAAAA	GGCGGGCATG	11760
TTTGGGTCAA	AATTAAGCAT	GTTAGAGAGA	TAGGGATCGC	GATAGGTACC	GTCATAGTTT	11820
TGGCAACAGT	TAATTACTTT	TTTCGCCTCA	GATAGCTCCT	CTTGGCTTAA	CTTGTCTTCT	11880
GCTTGAATCA	TATAGGTATC	CTCTACAAAC	CAGACGATCT	GTGACTGGCA	TCTTTAGCCT	11940
GCTCGAGTTT	ATTGACATAA	TACTCTCGTT	TTTCTTCGAC	TTCGTGAATG	ACAGGCTCAT	12000
CTTTCTTACC	ATGAAGACGG	ACAATCTTGG	CCGGAATACC	GACAACCGTC	ACGTCACTAG	12060
GTACATCTGC	TACGACAACT	GCTGCAGCAC	CGACCTTGGC	ATTTTCACCA	ATTTCCACAG	12120
GCCCCATAAC	TTGGGCATGG	GCTGATATGA	GGGCTCCCTT	TCGTACAGTC	GGATGGCGTT	12180
TGCCACAGTC	TTTCCCTGTT	CCCCCGAGAG	TCACTCCGTG	ATAGAGAAGA	ACGCCTTTTT	12240
CAACAATCGC	TGTCTCTCCA	ATCACCAGAC	CAGAACCATG	GTCAATAAAA	ACACCTGAAT	12300
CAATCTGGGC	TCCTGGATGA	ATCTCAATCT	GAGTCCAAAA	GCGCCAAAAA	TGACTGTACA	12360
TACGAGCTAA	TAGTTTGAAG	CCGTGCTTCC	AGAGAAAAATG	CGAGAGACGG	TGGGCCGCCA	12420
AGGCCTTGAC	ACCTGGATAA	GTCAGCAAAA	CCTCCAAAGT	GGTGCGGGCC	GCTGGATCAT	12480
TTTCTTTTAC	AATATCAATG	GTTTCGCGCC	ACCACCCCAT	ACATTTCTCC	TTTTCTTATT	12540

947

CTGAATCTTT TGATGTTTCT GTAAATTCTT TCTTAGGTTT GTAATCCTTT TGATGACGTG	12600
GGCGGTGAGG GCGCTCAGAC TTTTCACCTT TTTTCATCATG CTCAGGTTTT GGCGGACGAG	12660
GTAGAAGAGC CTTTCATAGAG GCATCGATAC GGCCTTTTTC ATCAATTTTG ATAACCTTAA	12720
CATCAACTTC ATCCCCGATT TCTACCAAAT CCTCTACACG ATTGGTACGA GTCCAAGCCA	12780
TCTCAGAGAT ATGAACAAGG GCATCTGTCT TATCAAAGAG GTTAACAAAG GCACCAAATT	12840
<hr/>	
TCTCGATACG AACGACTTTA GCACGGTAAA CTTTCATCCAC TTTGGCTTCA CGAACCAAAC	12900
CAGCAATAAT TTCTTTGGCA CGGTTAATAG CATCTTGGTC ACTAGAGTAG ATAGACACAT	12960
TTCTTCTTTC GTCTATATCA ATCTTAACAC CTGTTTCAGC GATAATCTTG TCGATGGTTT	13020
CTCCACCCTT ACCGATGACA ATCTTAATCT TGTCCACATC AATCTTGATC GTATCAATTT	13080
TGGGAGCAGT TGGAGCCAAT TCTGGACGAA CTTCTGGAAT GGTGCTTCA ATGACATCAA	13140
GGATTTCAAA ACGCGCTTTC TTGGCTTGAG CAAGAGCCTC CGTCAAGATT TCTGCAGTAA	13200
TCCCTTGAAT CTTGATATCC ATTTGAAGGG CTGTAATCCC ATCAGGAGTA CCTGCAACCT	13260
TGAAGTCCAT ATCTCAAAG TGATCTTCCA AACCTTGGAT ATCTGTCAAT ACTGTGTAGT	13320
TATTTCCATC TGAGATAAGC CCCATAGCAA TACCAGCTAC TGGCGCCTTG ATTGGCACAC	13380
CACCAGCCAT AAGGGCAAGA GTTCCCGCAC AGATAGAAGC TTGAGATGAA GAACCGTTTG	13440
ATTCCAAAAC TTCTGCTACT AGACGGATAG CGTAGGGGAA TTCTTCCAAG CTTGGCAAGA	13500
CTTGAGCAAG AGCACGCTCA CCAAGGGCAC CGTGACCGAT TTCACGACGA CCTGGCGCAC	13560
CGTAACGACC TGTTTCCCCT ACAGAATATT GAGGGAAGTT ATAGTGGTGC ATAAAGCGTT	13620
TCTTGTACTC TGGATCCAAA CCATCAATGA TTTGAGTTTC TCCCATCGGA GCCAAGGTCA	13680
AGACTGAAAG AGCTTGAGTT TGCCACGAG TAAAGAGACC TGAACCATGT ACACGAGGAA	13740
GGAAGTCAAC AACCGCATCC AAAGGACGGA TTTTCATCGAC CTTACGACCA TCAGGACGCA	13800
CCTTGTCTTC TGTAATTAAA CGTCGCACTT CTGCGTGTTC CATTTGTTCC AAGATTTGAG	13860
CCACATCAGC CATAATACGG TCAAATCTT CTGGTCCGC ATATTTTCT TCGTAAACGG	13920
CAGTCACTTG GTCTTCACT ACTTGAGTCG CAGCTTCAGG GGCCAATTC TCTTCTACTT	13980
GAACTGCCTT TTGGAGGTCA CTGTTGTAGG CTGCAATGAT TTCAGCTTGC AATTCAGCAT	14040
CCACGTGAAG CAATTCCACT TCTGCTTTTT CTTTACCGAC AGCAGCAACG ATTTCTTCTT	14100
GGAAGGCAAT CAATTCTTTG ACAGCTTCGT GCCCTTTAAG GAGCGCTTCC AACATGATTT	14160
CTTCTGACAA TTCTTTGGCA CCAGACTCTA CCATGTTGAT AGCGTGCTTG GTTCCAGCTA	14220
CTGTCAATTC AAGAAGAGAT TGCTCTGCTT GTTCTTGACT TGGGTTGATG ATGATTTGGC	14280

948

CATCTACATA TCCCACCTTGT ACCCCAGCAA TTGGTCCGTC AAATGGAATA TCTGAAATAG	14340
ACAGTGCCAA AGATGAACCA AACATAGCAG CCATTGGTGC AGATGCATTT TCATCATAAG	14400
AAAGCACTGT ATTGATGACT TGGACTTCAT TACGGAAACC TTCCGCAAAC ATAGGACGAA	14460
TCGGACGGTC AATCAAACGC GCTGTCAAGG TCGCATCTGT TGAAGGACGT CCTTCACGTT	14520
TCATAAAGCC ACCAGGAAAC TTCCCAGCCG CATACATTTT TTCTTCGTAG TTGACTTGGA	14580
GTGGGAAGAA ATCCCCAGTT GCCATTTTCT TAGACATAAC GGCAGCAGTC AAGACAGTTG	14640
ACTCACCGTA ACGTACGACA ACAGATCCAT TTGCTTGCTT AGCAACCTGA CCAGTCTCTA	14700
CAATTAATC ACGACCCGCA AAAGTCGTTT GAAACACTTG TTTTGCCATT TTAATCCCTT	14760
TTGGATTGAT GAAATTATAC GCCTTGCCTA CAAAGATCAA GATACCAAGG ACGTCAAAAG	14820
CAAAGTAAAA ATAGGAACT GACGAAGTCT TCGATGAAGA CAAGACAGTT TATCTTTTTT	14880
ACACAGCTTT TCGGCCGTGT TCAATTACAC AAGATATTTT GGACGGTTCG GCTTGCCGAA	14940
CATTTCTGTA GAAAAATAGG AAGGTGACGT CGCACTCGAC GAGTGCTAGG AAGCTTATCT	15000
TTTTTCCTAA GAAATGAGAC CAAAATTCAA GTCATCAAGA TACCAAGCCG TCAAGCAACT	15060
CAAAGGAAGA TAGGAAATCG AACGACGGAG CGACTACTCC TAGGGAGATT TATCTTTTTT	15120
CACAGAGTTG TAGGCAAGTT CAGTTTTCAA GATACATCAT TAGAAAGGTT TAATACTAAA	15180
GTATCTAAAG CTTTCACGCT AATCGCTATC GGGCGATTAG CTAATGCTT TACTAACTCT	15240
CTCGTCAAAT AACATCGATT TGACTCACTC GTGTCGTAA ATCTTACAGT TTAATGTCAT	15300
TGTATTATTT AATACCTTCA TCTTTGTATC AAGTACGTAC AGAATTTATT TTATCATATT	15360
TTTCTTAAAA AGTGAGGTCT TTACCATTAA AAAGGAACCA TTCCCCTCAC CTGAGAAGAA	15420
TGGTTTGCTT TTATTATCCT AGAGACTGGT GATTAAACAA GGCATGGGTT GCTTGATGGA	15480
TGTATTTTGC TGTATCAGCA TTATTCATCG TATAGAGATG CACACCGGCA ACATCCTGAG	15540
TTACCAAGTC CACGATTGCG TCCACTGCAT AGGCAAGTCC TGCTGCTCTG AGCGACTCAG	15600
GGTCATGCTC ATACTTGTCT AAGATGGCTT TAAATTTGCG TGGAAGATGG ATATTCTCAC	15660
AAGTCTTCAA GAGTCGGAGA GCCTGATTTC GATTCAGAAT TGGCATAATT CCTGCATGAA	15720
TGGGAACATC AATCCCAGCC AAGATACACT TGTCTGAAA ATCATAGAAG CGCTCATTGT	15780
CAAAGAAGAG CTGAGTTACG AGGCTCGAAC AGCCTGCATC CACTTTCTTC TTAAGATTTT	15840
GAATATCTGA AATCTGATTT GGCGAATCTG GATGCCCTTC TGGATAGCAA GCTCCAATAA	15900
TATCAAAGTG AGGGGTTTGT TCCTTGATAA ACTCAATCAA GTCGGTTGCA TAGCGGAAAT	15960
CCTTTTGTTG TTCCACGTCT GGAATAATAT CCCCACGAAG AGCCAAGATT TTCTGCACCC	16020
CAACTTTGTC CAAGTCAGCA ATAGTTTCAG CAACCTTGTC CTTAGTTAGA TAAATAGCTG	16080

949

GCAAGTGGGC AATGGTCGGA ATCGCCAAAT CATTTTGGAT AAAGTCAGCC AAACGAACCG	16140
TCGTTTCCTT GATATTAAAT TTATTATTGC TGGCAGTTAC ACTGATAAAA TGGGGAGCCA	16200
ACTCCTGCAT ATCCTGCAAG GCTGAAATAA TGTATCATT ACCCACGGCT GGGTTTGGAG	16260
GGAACACTTC AAATGAGAGT GACGGTGTTT GCGGTGACAT ATGTAATAAC CTTTCTAGT	16320
TGATTTCTTT TTGAACAACC ACTGTATGGA GAGAAATCCA ATCTTACAAT TTCTCACGGC	16380
<hr/>	
CAGCTTTAGC TGCTTCAACA AGGCGGATCA AGCTTTCTTT TGTTCCTGGG ATACCACGTG	16440
TTTTCAAACC ACAGTCAGGG TTGATCCAAA CTTTCTTGCT TGGCACTTTA GCAAGGATGG	16500
CTTCGATTGT GTTGTGATT TCGCCTTCAT TTGGTACACG AGGTGAGTGG ATATCGTAAA	16560
CCCCAGGTCC CACTTCTGTT TGGAAGTTT TCGCTTGAG TTCGTCCAAG ATTTCAAGGT	16620
TTGAACGGTT AGCTTCAAG GAAATAACGT CTGCATCCAT GTTATCGATA GCTGGGATGA	16680
TATCTGTAAA TTCTGAGTAA CACATGTGAG TGTGGATTG TGTGTCTGGC GCTACTGTTG	16740
AGTGATACAA GCGGAAGGCA GGAATAGCCC AGTCAAGGTA GTCTTCGTAC CAGTCGCTAC	16800
GGCGGAGTGG CAATTTTCA CGAAGAGCAG CCTCGTCGAT TTGGATGATT TTCACACCAG	16860
CAGCTTCAAG GTCAAGTACT TCATCCTGA TAGCAAGGGC GATTTGGAGA GTTGAATCCT	16920
TGATAGAGAT GTCTTCACGT GGAATGACC AGTTAAGGAT GGTAACAGGT CCAGTCAACA	16980
TACCTTTAAC AGGTTTGT TTACGACTTT GTGCATAGCT AGACCATTTA ACAGTGATAG	17040
GGTTAAGACG AGTGACATCA CCCCAGATGA TTGGTGGTTT TACCCACGC ATACCGTATG	17100
ATTGTACCCA TCCATTTTGA GAGAAGAGGT ATCCTGACAA GTTTTGACCG AAGTACTCAA	17160
CCATGTCATT ACGCTCAAAT TCACCGTGAA CAAGGACATC AAAGTCAATA TCTTCTTGCC	17220
ACTTGATCCA TTCGTCAATC GTTTCAGCAA GGAAAGCGTC GTACTCTTTT TGAGACAATT	17280
CACCTTTACG GTAAGCCAAA CGTTTGGCAC GAACTTCTTT TGTTCGAGG AATGAACCAA	17340
TCGTTGTTGT TGGAAGAGCT GGAAGTTGA AAGCTTCTTC TTGGATAGCT TCACGTTCTG	17400
CAAAGGCTGG CAAACGAGTG TAGTCTGCGT CTGTCAAGCC AGCGATACGC GCACGAAGTT	17460
CAGCATTTTC ACCAACACGC TCAGTCGCAA AGAGTTCTTT GTTGGCTGCA AGAGCTTCTG	17520
AACCTTGACC ATTTCCGATA GCATCCAAGT CACGGATTTC ATCCAATTTT TCAACTGCAA	17580
AGGCAAAGTG GTTCAAGAGT GCTGGTTCAA ATTCTTCATT AGCAGTTGTA AATGGCACAT	17640
GAAGAAGTGA GCAAGAGCTT GTCAAAACGA TGTTTTCAGC TGGAAATTTG TCAAGAACAG	17700
CCAAGCTCTT TTCGTAGTTG TTGCGCCAGA TGTTTTTACC ATTGACAATA CCTACATAGA	17760
GAGTCTTGTC AGCTGGGAAG CCACCTTTAA CGAGTTCAAG AGTTTCTTA CCTTCAACAA	17820

950

AGTCAAGACC GATAGCATCT ACTGGTAAGT TTACAAGGTC AGCGTATACG TCACGAACAT	17880
CACCGAAATA AGTTTGAAGC AAGACTTCAA GACCTTTTTT GTCAGCCAAG AGTTTGTGTG	17940
AAAGGTTCAA GAAGAGAGCT TTTTCTTCAG CTGTCAAGTC TTTTACAAGA GCCGCTTCAT	18000
CCAATTGGAT GCGAGTCGCA CCAAGTTCAG CCAATTTAGC AAAAATTCTT TGGTAAGCAG	18060
CCACTAAGCT ATCTACGAAG TCGTCTGCTT TCACGCCTTC TTCAAAGTCT GACAATTGAA	18120
GGAAAGTGAA GGGACCTACA AGAACAGGAC GAGTGTTCOA TCCAAGTTCT TTGGCTTCTT	18180
GGAACTCATC GAAAATCTTG TGACCAGCCA ATTTTACTTG AGTGTCTTTT TCAAATTTAG	18240
GAACGATGTA GTGGTAGTTA GTGTTGAACC ATTTCTTCAT TGGAAGGGCG CGAACGTCCC	18300
CTTTTCTCC CTGGTAACCA CGTCCCAAAG CGAAGTAGCG CTCAAGGTCA GACAAGTCCA	18360
AGTTTGAAC GGATGCAGGC ACCACGTTGA AAAGGAAAGC CGCATCTAGG AAGTTATCAT	18420
AGTGAGAAAA GTCATTTGAT GGAATTTGAG TGATGCCTTT TTCTTTGACA ATGTTCCAGT	18480
GTTTAGCACG CAAGTCTTTT GCTGCTGCTA AAAGTCTTTC TTCTGAGATT TCTTTTCTAA	18540
AGTATTTTTC AGTTGTAAAT TTAAATTCAC GGAATTCGCC CAAACGAGGG AAACCGATGA	18600
TTGTAGTTGA CATGATGTGT CCTCCAAAAT TTGTTGTTGA AACTATCTTA ACAGAAAAGA	18660
AAGCGTCTGT ATAATTGTAA AAAATTAGGG TTTGATATAG TTTGAAACTA TATATCTGTT	18720
TCGGACAAAA GAAAAAGACT TGAAGCAAAC GTCTCAAATC CTTTGTAATT CTTACTTTAC	18780
AGCTATATTC CAATTAGAAT ACTAAAACAT GTTATTAGTA ATTCTTATAA GTGACTATGA	18840
CCTGTTATTA GAAAAGACTA TAACTGATTC TAGTCAACTT TTTCCCTGTT CAAGTGGGAC	18900
GATTGCTAGT GTCTTTCTTA AACTGGCTAG GACTTTTAAG ACTGTATCCA ACTGAGGACT	18960
AGTCTTTCCT GTCTCCATCC TAGCTATGAC AGGCTGGCTT ATTCCACTGA CTTCTTCCAG	19020
CTTTTCTGA CTGATTCCTT GTTCATACCT AGCCTCAATC AACTCGCTCA TGATAGCCAC	19080
TCGCATATCA CTMTCAAGGA TTTCTCCTT GCTAAAGAGC TCAGATGGAC ATCCTTCCAA	19140
TTACTCCCAA TAGCACTATT CTTCACTACT TAACCTCTT TTTTTTACGT CTATGTATTT	19200
TTAAAAAAT GAGCGAATTA TGATTGATA GATTGACCAG TGGGTTTAAA GTTGGTGCTA	19260
GCCTATTCTT TAAGCGATTT TCCTTTTCTA GGATAAAGCA GTTCCTGCTT GCTTAACCCC	19320
AATTTTCCAC GATGAATCCA ATAGTAAATG GTTGAAATTC CCACGTTAAC CCCTTTAGCC	19380
ATCACCATCA TTPCAGGCGA AAATTTTGG TTATGTTTTT GGTATGTAT AGTGGAGAAT	19440
CTTTTCTTT AGTTTCTTAA GACTGTTGAG CGTAGTCGGC AGAATAAATC TCTTTGAAGC	19500
GCCCTTTTCC AAGACATTGT CGGACTGTCC CACGCTTGAT TTCAGTGTGG ATAGTTTGAG	19560
GAGCTTTTCC AAGTAGAGAG GCAATTTCTC TATTGATTT TCCTTCTTTT TTCCATCGTT	19620

CGATTAAGCG ACGGCTATCG ATTGTCAAAT GTTTCGCTTT TGTAGTATAA TTGCTTGCA	19680
TTTCTGTGCC TTTTAATCAT TTCAATCTTA AATTGGACTT TTTTACTTG GGTGTACTT	19740
AATCTATGAG GAAGACAAGA AAAAGAATAT CAATCAAGTA AAGTCACAAA GTCACATTAG	19800
CTCCGAGCAA CCATTGCAAA TTGAGGTACT CACACAATGA TTA AACATT TCTCTCGCC	19860
CTTTCGGTCA TTCTCTTTTC TATCCCTATC ATA ACTTATT CTTTTTCCC ATCTTCTAAT	19920
CTTAACATTT GGCTATCTAC CCAACCTATC TTGGCACAGA TTTATGCCTT CCCCTTAGCT	19980
ACTGCAACTA TGGCTGCTAT TTTAAGTTTC TTATTTT TTTT TCCTATCTTT TTACAAGAAA	20040
AATAACAAA TACGGTTTTA CTCTGGCATT TTGCTCTTAC TATCGCTCAT ATTACTATTA	20100
TTCGGAACAG ATAAAACCTT TTCTTCTGCA TCAATAAGA CTA AACCTT AAAATTAGTA	20160
ACTTGGAACG TCGCTAATCA AATAGAAGCA CAACATATTG AGCGAATTTT TAGCCATTTT	20220
GACGCCGATA TGGCTATATT CCCTGAACTA GCTACCAATA TCAGAGGTGA GCAAGAAAAC	20280
CAGAGAAATCA AACTATTGTT TCATCAAGTT GGACTTTCTA TGGCCAACTA TGATATTTTC	20340
ACTTCTCCAC CTACCAATAG TGAATAGCT CCTGTGACTG TGATTGTCAA GAAAAGTTAT	20400
GGTTTCTATA CAGAAGCTAA AACTTTTCAT ACAACACGGT TCGGGACAAT TGTATTACAT	20460
TCGAGAAAAC AAAATATACC AGATATCATT GCCTTGCATA CTGCGCCTCC TCTGCCAGGT	20520
TTAATGGAAA TCTGGAAGCA AGACTTAAAC ATCATTGATA ATCAATTGGC TTCAAATAT	20580
CCAAAGGCTA TTATTGCAGG TGATTTTAAT GCAACTATGC GTCATGGAGC ACTTGCAAAA	20640
ATAAGCTCTC ATAGGGACGC ATTAATGCA CTGCCACCTT TTGAAAGAGG AACTTGAAT	20700
AGCCAAAGTC CAAACTTTT TAATGCAACA ATAGATCATA TTTTATTGCC TAAAAACCAC	20760
TACTATGTTA AAGATTAGA CATTGTAAGT TTTCAAACT CTGATCATAG ATGTATTTTT	20820
ACAGAAATCA CATTTTAATT ATTTTATATA AAATCACCCC TCTAATGTTT ATAACTAGA	20880
GGGGGAATTT GTATCCTACT ATCGTTTAAAC GCACTTCTGC ATTGACTTTT TCTTCGAGAG	20940
ACGCTTGGAT TTTTCCATA TAGCGTGCGA CTTCTTCGTC CGTTAAGCTG TCTTCTGGAT	21000
TTTGAAGGT CAAGCTATAA GCCATTGACT TCATACCAAG TCCCAGTTTT TCACCTGAGA	21060
AGACGTCAA GAGTTTGATA TCTGTCAAAC GTTTCACGCC GGCAGCTTGG ATAGCATCTA	21120
CAACTTCTTG GTGAGTCACT TCTGCCTTGA GGAGAAGGGC AACGTCACGG CTGACTGCTG	21180
GGAATTTGGT GATTTCCACA AATGGAACAG CAGGTTGGAG CGCCCCCTCG ATGGCTGAAA	21240
GGTTAAGCTC AGCTACATAC GTTCTGGA TATCGTAAGC CTGGCAGTG ACTGGATGCA	21300
CTTGCCAAAG GAAACCAAGA ACTTGGTCAC CGAGTGAAAT CACGGCTGTA CGACCTGGAT	21360

952

GAAGGCTAAC GATTTCAGAT GTTGCTGTAT AGGTTACTTG GAGTCCCAA CGAGTAAATA	21420
GGGCTTCAAG GATTCCCTTA GCATAGAAGA AATCAACTGG AACTGCTGCT GTTTGGAAAT	21480
CTTTTTCAGC AACCAAGCCT GTCAAGGCAA AGGCAAAGCT GTTGATCTCA TTTGGAAGTT	21540
CTTCTTTTGG ATTACCTGTT TGTCAAAGA CTTTCCCAAT CTCATAAAGG GCCAAGTTTT	21600
TATTCTTACG AGCCACGTTG TAGGCAACGG TATCAAGGAT CCCTGAAATC ATATTTTGAC	21660
GGAGGACTGA ACGATCCACA GTCATTGGCC ACATGAGTTC AGTAAGGTTA CTTGGTTGAG	21720
CTGTGAACTC AACTGCTTTT TCAGGAGTTG TCAGAGCATA GGTGATGATT TCTGTCAAAC	21780
CTGCTCCTTC AGCAATGGTA CGAACTTGAC GCGGAGTTT TTGTATCACA GTCAATTCAC	21840
CAGCTGTACC ATCGTCTTTT GGAAGGCTGG TTGGCAAGCG GTCATATCCA TAGATACGAG	21900
CGATTTCTTC AAAGAGATCA GCTTCGATTG TGATATCCCA ACGACGACGT GGTACGCTGA	21960
CTGTAAAGCT GTCTGCATTT CCAGAAAGAC CAAAGCCAAG ACGACGGAAG ACGTCTTCTA	22020
CATCAGCATA AGACAGCTCA GTTCCGAGGA CACGGTTAAC ATCAGCAAGG GTTGAAGAAA	22080
CTTCCACATC AGAGGTATCA AGCTCACCCG CTGAAACGAT ACCCTTACGC ACCGTCGCGC	22140
CTGCAAGCTC TGCAATCATG CTAGCTGCCG CATCAAGGGC TTCATTAACT GTTGCCACAT	22200
TAATTCCTTT TTCAAAGCGA GAAGATGACT CAGAACGAAG GTTCAGGCGA CCACTTGTCT	22260
TACGGATAGA TTTGCCATTA AAAACAGCAG CTTCAAGGAT AACACGACTA GATTTTTCAG	22320
AAATTTCTGT AGCCTGACCA CCCATAACAC CGGCAAGGGC TACTGGTTTG TCAGCAACTG	22380
TAATCACGAG GTCTGTCTCA GCCAAGTCTC GTTCTTCACC GTCCAGGGTC ACTAATTTTT	22440
CACCATCACG CGCTTCACGC ACACGGATGT CAGTCCCTTC AAATGTGTCC AAGTCAAAAG	22500
CATGCATAGG TTGACCAAAG TAGAGCAGGA TGTAGTTTGT CACGTCTACA ACGTTATTGA	22560
TGGGACGGAT GCCTTCGTTT ATGAGAAGGT TTTGCAACCA TTGTGGACTT GGTGCGATAG	22620
TCACATTGTC CAAGATACGA GCTGCATAGT AAGGCGCCTT GTCTGTCTCA ATGCTGACAG	22680
AAAGGGCATC TGCCGCAGCT TCATTAGTTT CTGTTAGAGT AAATTTTTTA AAGTTGACTG	22740
CCTTGTCTATA GATGGCTGCC ACTTCGTGAG CCACTCCACA CATAGAAAGG GCATCTGCAC	22800
GGTTTGGTGT GATGGAAAGT TCGATGATTT CATCATCCAA GTCTAGGTAA GAAAAGACTT	22860
CCTCACCTGG CACGGCATCT TCAGGCAAGA TTTGGATGCC ATCTGCGAAT TCCTTAGGCA	22920
CAACTGAGTC AGAAATTCCT AATTCACCAA GTGAACAGAT CATTCCAAGT GACTCCAAAC	22980
CACGGATTTT TCCTTTTGTG ATTTTGTAGT TATCAGCGAT ACGAGCTCCT GGAAGAGCCA	23040
CCATGACCTT GATCCCAGCA CGCACATTTG GGGCACCACA AACGATCTGA CGCTCTTCTT	23100
CTTCGCCAAC GTTAATCTGA CAAACATGGA GGTGAGTCTC TGGCACATCT TCGCAAGACA	23160



AGACCTCACC GACGACAATT TTTGAGAGAC CAGCAGCTGG TGATTGACA CCCTCTACCT	23220
CGATCCCTGT AGTTGACATT TTTTCAGCCA ACTCTTGTA TGGCACATCA ATGTCCACCA	23280
ATTCTTTTAA CCATTTATAA GATACAAGCA TAATTTAGTT CTCCAGAATG ACAGTTGTCA	23340
CTCTAGTTCT TTTCTTTCC TATCATTTCA ATAGAAGAAT CCTCTTCTTA CCTTAATTC	23400
TTTCTCAGTA ACCAATCCGT ATCTACTTTT TGACCAACCA TAAAATGATG TTGGCTAAAT	23460
TTTTCAAAAC CATATCGGTT ATAAAACGCT TGAGCTTTTG TATTATCCTC CCAAACACCT	23520
AGCCAAGCCC AAGAAAACT ATTTTTTGTA GCAAGTTCAA GTGCGAATTC AAACAGTTGC	23580
TTACCTAGTC CAAATCCTTG GAATTTTTGT AGCACATAGA GACGTTGAAT TTCAAAAGCG	23640
TCCTCTAATT CTCTCTCAGT TTGAGCACTT CCCCAGTTGA CTTTGAGAAA ACCAGCTATC	23700
TCCTCCTCAT GCATAATGAA ATAGGTTTCA GAGTCAGGAT TTCCCAACTC AGTTGACAAA	23760
GTTTTAGAC TATAAGCCTC TTCAAAGTAT TCCTGTAAGT GCTCTTCCGT ATTATCATAC	23820
GCAAAGGTTT CACGAAAGGT TTGTTTGGA ATTTTAGCCA ACACCTCAAC ATCTGCCATT	23880
TCTACTTTTC TAATCATAT TTAAGTGT CTGAGAAGCG GACATCTCCT TGGTAGAATC	23940
CACGGATATC GTTGATTCCA TAACGGAGCA TAGCTACAG CTCTTGTTCA AGACCAAAGG	24000
CAAAGCCAGA GTATACAGTC GCATCGATAC CACTCATTTT AAGGACACGT GGGTGAACCA	24060
TACCGGCCCC CATAATTTTG ATCCAACCTG TTTTCTTACA TACATTACAG CCTTCTCCAC	24120
CACACTTGAA GCAAGAAACA TCCACCTCAA CAGATGGCTC TGTGAATGGG AAGTAAGATG	24180
GACGCAAACG AATTTGACGC TCTTCACCAA ACATTTTTTG GACAATCAAC TGAAGCGTTC	24240
CTTGAAGATC AGCCATAGAG ATATTTTTCC CAACTACCAA GCCTTCGATT TGGTGAAT	24300
GGTGACTGTG GGTCGCATCG TCCGTATCGC GACGGAAGAC ACGCCCTGGC GAGATCATCT	24360
TCAAAGGACC TTTAGAAAA TCATGGGCAT CCATAGCACG CGCCTGAACT GGAGACGTGT	24420
GGGTACGGAG CAAGATTTCT TCAGTGATAT AGAAAGTATC CTGCATATCA CGAGCTGGGT	24480
GGTCTTTTGG AAGGTTTATA CGTTCAAAGT TATAGTAGTC TTGCTCCACT TCAAAACCAT	24540
CCACGACTTG ATAACCCATA CCGATGAAGA TATCTTCGAT TTCTTCACTG GTTTGTGTCA	24600
AAACGTGACG GTGACCAGTC GCAACTGGAC GACCTGGAAG CGTCACATCT ATACTCTCGC	24660
TAGCCAGTTG AGCCGCGACT TTCTTTTCTT CCAAGAGCTT AGCTGTTTCT TCAAAAGCAG	24720
CAGTCAAGAC ATCAGGAGCT TCATTGACGT GTTTCCCGAT GATTGGACGC ATCTCAGCAG	24780
AAACATCTTT CATCCCTTTG AGGATTTTCAG TGAGCGAACC CTTTTTACCA AGGACAGAGA	24840
CACGCAAATC TTGCATCTCT TTTTCATTTT CAGCAGTAAT CTGCTTCAAG CTAGCCAGCG	24900

954

TTTCTTCGCG AAGCGCTTTT AATTGTTCTT CAATAGTTGA CATATTCCT CCATCAGTCT	24960
CTCGTAGATA AAAAGAAAAC CACATGCCAA AAACCTCCACT CGGAGCGTTG ACACGCGGTA	25020
CCATCCGTTT TCATCTGACA AGTCAGACCT TCATTCTAA ATCCATGCGC AAGTGAATTC	25080
ACCCAGCTTT CATATAGAGA GCTTGCACTC ACGGCTCTCC TCCCTGATAT ACTTCCCTTG	25140
GGCTACTAGT CTTTCAGATT CCTATTCAAT TACTACTTAG TTTATCAGAT TTTTACCATT	25200
CTTGCAAGAC CTATCTTACT TCTGCTTGTT AGCTTATTCT TATCTAAAT TATATAAACc	25260
TTATCTAAAT TAACTATTTA TAATTTTGT AACAAAATTA AATTAATTGA CACTCCCCTA	25320
TAAATAAAG AAGTTTAGAA TTTAATGTCT TCCAACTTC TTTATTCCAT ATTTAATGAA	25380
ATGCCACCTT AACCGTGATA ATAGCTAGTC ATCAATAAAA AACTATTTGA ATAAGGATTC	25440
TCCATTTGAT TCAATCACTT CTTTATACCA AGTAAAAGAC ATTTCTTAT ATCGATTTAA	25500
TGTACCACTT CCATCATCGT TTCGATCAAC ATAAATGAGA CCGTACCTTT TAGAAAGTTG	25560
TGCAGTGGAC ATAGAAACAC AGTCAATACA TCCCCAAGAC GTATAGCCCA TAATTTCAAC	25620
ACCATCCTGT AGAGCTTCAG CAACTTGCAA TAAATGTTCT TTCATATACT GAATTCTATA	25680
ATCATCTTGG ACGGTTAAGT TATTAAGTTC ATCTTTTATT AGTNGATCTT TAGCACCTAA	25740
TCCATTTTCT ACTATAAATA ATGGGATTG ATAACGGTCA TAATATCTAT TTAATAATTAT	25800
ACGTAGTCCA ATTGGATCAA TTTGCCATCC CCACTCTGAA GACTCTAAAT AAGGATTTAC	25860
TAAACCACCA ATAATATTCC CTTCTCCTGA ATTATACTGT GTTGAAGAG CAGATTGAGT	25920
CACACTCATG TAATAGCTAA AGGATAAAAA ATCTACGGTA TAATTTTTTA ATAACCTGTC	25980
ATCTTCAGCT GCAAACCTTA TGTTAATGTC ATTTTCCTTA AAATATCTTT TTGCATAATT	26040
CGGATAATAA CCTCTAACAT GCACATCTGA AAATAGATAA TTTAGATTCT CATACTCATG	26100
AGTCGCCCCAT ACATCTTTTG GATTGGAGT CATTTGGATAA GCTGGCATAG CTAATACCAT	26160
ACATCCCACC TTAAACTCTG AATTAATCTC ACGAGCAATT TTTGTAACCA AACTTGAGGC	26220
GACTAATTCA TGATGTATAG CTTGATATAA TTCTTGTTTC GAAAGATTCT CCTTAGGTAT	26280
ATCTATTCCT CCACTAGTAA ATGTAATTC CAAAACAGAG TTTACTTCGT TAAATGTAAG	26340
CCAAATTTTA ACTTTATCTT TATACCTTTC TAAAACGTGT CGAGCAAATT TTTCAATAAA	26400
ATGAATCATT CTCCTATCAA CCCATCCATG ATATTTTCTT GCTAAATATA ATGGAGTCTC	26460
ATAGTGTGAA AGAGTTACAA GTGGTTCTAT CCCGTGAGCA TGTAGTTCAT CAAACAATTC	26520
ATCATAATAT TTCAACCCAG CTTGTTAGG TTCTTCCTCA TCTCCTTTTG GAAAAATTCT	26580
ACTCCATGCA ATAGAAGTAC GAAAAACATT AAAGCCCAT TCAGAAAACA AGGATATATC	26640
TTCTTTATAT TTATGATAAA AATCAATACC TATCAATTTT AAGTTATCTT CTGTAGGATT	26700

955

TTCTGTTGCT TCTCCTAATC CACCTTTGGG TAACACATCC TGAAGTGATA AGCCCTTACC	26760
ATCTTCATTA TATGCTCCCT CTAATTGATT AGCTGCAACA GCTCCACCCC AAAGAAAATC	26820
ATCTGGAAAA ATGGTCATAA CTTTCTCTCA TTATAATATT ACCAGTAATT CCTTAGAATG	26880
CTCGATTGTC TGATTATTAG GTAATACTAA TACATCTAGA AAATCATTGG TATTCGTTAC	26940
AATTACTGGT GTAAGTGTTC CGTAGCCTTT AGTCTTGATT AAATTCAAGT CCATTTCAAA	27000
AATCAACTGA TTTTGGAAAA CTCTGTCTCC TTCTTCTACA TGAAGTAATA AACCTTGACC	27060
TTTGTAGCTCA ACAGTATCTA ATCCAATATG AATTAGTAAC TCAACACCCT CATCACTCTT	27120
CAATCCAATT GCGTGCTTAG TCGGAAAAAT ATTTGTAATT TTCCCATCAA ATGGTGATA	27180
AACCTTACCT TCATTGCGA TAATCGCTAC TCCGTCTCCA ATTAGTTTAT CTGAAAATGT	27240
TTTATCCTGG ACATCGCTTA ACGGAATGAT TTCTCTGAT ATAGGAGAAA ATATCATTTT	27300
TTTATTTGAA ACTCCAGCTT CAACTTCTAA ATTGCTAGAA CTCTCTTCTT CATCGATTCC	27360
AAATATATAA GCTAATACAA AGGTAATAAC AACCGAAATG ACCGCCACAA TTAAAGCATT	27420
TACAATATTT GATGGCACAT CAGAATAAAT AAATTGAGGC AACGCTATCA AAGATGGGAC	27480
AGCAAATAGA TATGCTTTAA CACTAGTAAG ACCTGCAAAT AATCCCGCTA ATCCACCACC	27540
AATCATAGCT GCATAAAGCG GTTTTTTATA TTTTAAAGTC ACACCATATA ATGCAGGTTT	27600
GGTAATCCCT GCAAGTAAGG CTGAGAAACC TGCTGCAAAA GCAATTTGTT TTGTATTATT	27660
ATTTTACTC TTTAATGCAA CAGCCATCGA AGCAGCCCCT TGAGCTAAGT TTGACCCTAA	27720
CATTGCTGGA AGAATTAATA CGTCTGGAGT AGCAATAGAT GCCGCCAAAA AAATAGGTGC	27780
AAAAGCCCAA TGCAATCCAG TCATAACAAT AAATGGCATA ATAGCACCAA GAATAGCTAA	27840
TGTAAGCCAT CCAGCTACAC CATACATTG CCCAACTAGA TTTGATAATC CTTACCAAC	27900
AATTACTCCA ATAGGTCCGA CTACAACATA GGCAATACAG CTTGATACTA ATAATACTAG	27960
CGTAGGTTGC AAAAACTCT TAGTAATAGC TAGTGTTAAT TTAGCAATTA TTTTTTCAAT	28020
ATATTTTATC AACCAAACCA TAATAAGAAT TGAACGACT GATGAACCAT AACTAGCTGG	28080
TGTCACAGGT GCACCAATA AACTAAGAGG ATTCCCTGAT TGCACCATTT GAACAAAATT	28140
TGGATGGAGA AGTACACCTG CTACAGACAT AGCTAATGTA GATGTTACTT TTAATTTTTG	28200
TGATGCAGAA TAAGCTAATA ACAGCGGTAA GAAATAATAT GGAGCATCCC CAAAAAATGT	28260
CAAAAAAGCA ATAGTCTGAG AATCTGATTG CAATATACCA AGCATTGGTA AAATGATTAC	28320
CAAGACTTTC AACATACCTC CCCCTAACAT TGCTGGAATG ATTGGAGTCA TGAACACGAC	28380
GATATACTCA ATGATTCTTT CTAAAATATT CCCTTTGTGC CCTTGAACAA CTGAATCGGA	28440

956

TTCAAAATTG CCAAGTTTAA CGAATTCCTT ATAATAATTA GCTACATCAT TACCAAGTAT	28500
AATTTGATAT TGTCCATTCT TTTTCATAAT ACCTATTACA CCTGGTATCT TCTTCACATC	28560
ATCATCATTG ACTAAATTTT CATCTTTTAA TTCTAATCTT AAACGTGTTA CACAATGGGT	28620
AACTCTATTG ACATTTTTTT CACCTCCAAT TACATCGAGG ATTTTTTGTA CCGTATCTTT	28680
ATAACTCATG GTATTCTCCT ATTCTATTAA TCTAAATTTT TTGTTAAGCG ACGAATATGA	28740
GCCATCAAAT AACTAATTC ACTAGAAGTC AGCAAATAAT TGTACTCCGT TTGTATAAAC	28800
ATTGCTACCT GTTCACCACA TTCATATTCT CTAGGATATT TATTTTTCAT TAATGCTAAC	28860
AAGTCTTCAT CATCATCGTC GG	28882

## (2) INFORMATION FOR SEQ ID NO: 141:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12835 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

GCCTATGTCT TTTTCAAAAA AATGCTTGAC TTGAGACGGG AACTAGGGAA GTCTAAAGGC	60
GGAAGGCATT GATTTATACT CTTGAAAAT CTCTTCAAAC CACGTCAACG TCGCCTTGGA	120
TTATATATGT AACTGACTTC GTCGATGCTT ATCTACAACC TCAAAGCAGT GCTTTGAGCA	180
ACTTGCGGCT AGTTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATTATA TTACTTTCTA	240
TTGTAGGAG GTGGCTTATG AAGATTCCTC TCTTAACCTT TGCAAGGCAT AAATTTGTTT	300
ATGTCTTGCT TACTTTGCTT TTTCTTGCTT TGGTTTATCG TGATGTTTTG ATGACTTATT	360
TCTTTTTTGA TATTCATGCG CCCGATCTAG CTAAATTCGA TGGACAAGCA ATTAATAATG	420
ACTTATTAAA ATCAGCATTA GATTTTCGTA TTCTCCAGTT CAATCTAGGT TTTTATCAAT	480
CATTTATTAT TCCAATCATC ATTGTTTTGC TAGGTTTTCA ATATATTGAG CTGAAAAATA	540
AAGTTTTACG ATTGAGTATT GGAAGAGAAG TGAGTTATCA AGGGTTAAAA AGAAAGTTGA	600
CTTGCAAGT TGCAAGTATC CCTTGTTTGA TATATTTAGT GACTGTGCTG ATAATTGCAA	660
TTATAACCTA TTTCTTTGGG ACTTTTTCTC CTCTTGGATG GAATTCTCTA TTTTCTGATG	720
GAAGTGGTTT ACAAAGACTC CTAGATGGAG AGATAAAAAG CTATTTGTTC TTTACTTGTG	780
TCCTACTAAT CGGTATTTTC ATCAATGCAA TCTATTTTTT ACAAATAGTT GATTATGTGG	840
GGAATGTGAC TCGTTCGGCA ATCACCTATT TGATGTTTCT TTGGCTTGGT TCTATGCTGC	900
TTTATAGTGC CTGTCCTTAC TATATGGTTC CTATGACGAG TTTGATGCAA GCTAGCTATG	960

957

GGGATGTAAG TTTGATGAAA CTCTTTACTC CTTATATCCT TTATATTGTC CCTTACATGG	1020
TGCTTGAAAA ATATGAAGAT AATGTTTAAAG AATTTTAAACA ATATTTTGCT AAATAGAAAAG	1080
ATTGTTTTAC TACTTCGTAT AGTTCTGATG ATGATTTTGA TAAACCATCT ATTGTCAACA	1140
GCGGTTCAAA AGCAGGATGC TGTATCTTT TTCAAGAGAG AATTGATTTC AATTTTTC	1200
TATAATGACT ATTCTGAAGC GAATTTAGAA ATCCCCAAAC TATTGTTAAA CCTTTCGCTT	1260
TTTCATGGTAG GATGGCTCTC TGTCATTTTA CTTGAAAGTG ATTTGGCAGA CCATTACCAT	1320
CACTTGATTC GCTATCAATC AAGCTCCTTT TTCGATTATA CAAGGAAACG ATTGGTTGTC	1380
ATTTCTAAAT TTTTACTCA AGATTGTTT GTCTGGTTTC TTGGTTTACT TCCTCTAGGA	1440
ATTCATTTCA AAACAGTCGC ACTTTTCTTT TTAATTGCTC AGTTAATGAT GTTGACTTA	1500
CTACTGTCTT ATCTGATAGC ACTGATTAGT GCGGGCGCTG GTTTTTCCTT TTTTCTCTAT	1560
TTTTTAGCAT TTGTGGGACA AGAATGGATG ATGGATCATA TTGTAACAGT GTATTAGTA	1620
CTCTTAAGTT TATTAGTTAT GTTGATTGTT AGTCGCTTGG AAGAGAAATT TAAGAAAGGA	1680
TAAACGATGA GACTTGAAAT TATAAATGGA CAGAAAATTT ATGGGAAAAG ACCTATTTTA	1740
AATCAGTTGA ATTTGGTGTT TCAATCAGGA AAAATTTATG GACTTAAAGG TGATAATGGA	1800
TCTGGCAAGA CGGTTCTTTT AAAGATACTT GCTGGTTATA TTAAGCTTGA CAAAGGAAAA	1860
GTCTTCAAG ATGGTAAAGT TTACGGGGTA AAAAAATCATT ATATTCAGGA TGCAGGAATT	1920
TTAATTGAAA AAGTCGAGTT TTTATCTCAT TTATCCCTGA GAGAAAATTT GGAAGCTGTTA	1980
AGGTATTTTT CATCTAAAGT TACGGAAAAA AGAATTGCCT ATTGGATTCA ATACTATGAT	2040
TTACAGGAAT TTGAAGACAT TGAATACCGT CATTTATCCT TAGGAACAAA GCAAAAAATG	2100
GCCTTGATTC AAGCCTTTAT TTCTCTCCT TCTATACTCT TTCTCGATGA ACCTATGAAT	2160
GCTTTGGATG AGAAGAGTGT GAGGTAAACC AAACAGGTCA TTTTATCTTA CCTGAAAAAA	2220
GAAAATGGTC TGGTTATCCT GACGTCGCAC ATATCGGAAG ATATTTTACA CCTTTGTACA	2280
GATGTATTAG TTGTCGAAAA TGGACATATA CAAATGTAAA GGATATACAA TCCTAGGAGA	2340
TGGCTTATGG CACATCTAAA ATCATTATT ACACGATATT CCAAGGTTTA TATTGGTTTA	2400
GTTCGCTGA TCTGGCTGTC TTTCTCTTT ATCCCTTGGG ATAAACCACT TCTGGGGATA	2460
AGGATTGACA TCTTCATCAT ACAGAAAATC TTGCTAGCTT TTGGAATTCT GTCCATTCTC	2520
ATGGCCTTGC TGTCCAAGAA AGTCAGTCTC TTTGTTTTTG GACTGATTTG CTGTCTTTCT	2580
CTTTGGATTA ACTTATTTAT CACATTTGCC ATTTTGCCGA TTTTGGCAA TTAACAGTC	2640
ATAAAAGTCG GAGAGGTTAG CTTGAAAAC AACCTCTTT TCCTTTTCAA AATGGGGATT	2700

958

CTTCCTTGAA AATAATCAGT AATTGTGCTA AAATTAAAGG AACATTCTAA AATATTCCGA	2760
ATTTAAAGTA AGGAAAAACA TGGCTAATAT TTTAAAAACA ATTATCGAAA ATGATAAAGG	2820
AGAAATCCGT CGTCTGGAAA AGATGGCTGA CAAGGTTTTC AAATACGAAG ACCAAATGGC	2880
TGCTTTGACT GACGACCAAC TAAAAGCAAA AACAGTTGAA TTTAAGGAAC GTTATCAAAA	2940
TGGAGAATCA CTGGATTCAT TGCTTTACGA AGCATTTCGG GTTGTCCGTG AAGGTGCCAA	3000
ACGTGTCCTA GGTCTCTTCC CTTATAAGGT TCAGGTCATG GGGGGGATTG TTCTTCACCA	3060
TGGTGACGTG CCAGAGATGC GTACAGGGGA AGGGAAAACC TTGACTGCGA CCATGCCGGT	3120
ATACCTCAAT GCCCTTTCAG GTAAAGGGGT TCACGTAGTT ACGGTTAATG AATACCTGTC	3180
AGAACGTGAC GCGACTGAGA TGGGTGAATT GTACTCTTGG CTTGGTTTGT CAGTAGGGAT	3240
TAACCTGGCT ACCAAATCTC CAATGGAGAA AAAAGAAGCC TATGAGTGTG ATATTACTTA	3300
CTCAACTAAC TCAGAAATCG GATTTGACTA CCTTCGTGAC AACATGGTCG TTCGCGCCGA	3360
AAACATGGTA CAACGTCCGC TTAACATATG CTTGGTCGAT GAGGTGACT CTATCTTGAT	3420
TGACGAGGCT CGTACACCTT TGATTGTATC AGGTGCCAAT GCGGTTGAAA CCAGTCAGTT	3480
GTATCACATG GCAGACCACT ATGTAAATC TTTGAACAAA GATGACTACA TCATCGATGT	3540
GCAGTCTAAG ACTATTGGTT TGTCTGATTC AGGGATTGAC AGGGCTGAAA GCTACTTCAA	3600
ACTTGAAAAC CTCTATGACA TCGAAAACGT GGCTTTGACT CACTTTATCG ATAACGCCCT	3660
TCGTGCCAAC TACATCATGC TTCTCGATAT TGACTATGTG GTGAGCGAAG AGCAAGAAAT	3720
CTTGATTGTC GACCAATTTA CAGGTCGTAC CATGGAAGGT CGTCGTTATT CTGATGGATT	3780
GCACCAAGCT ATTGAAGCCA AAGAAGGTGT GCCAATCCAG GATGAAACCA AGACATCTGC	3840
CTCAATCACG TACCAAAACC TCTTCCGTAT GTACAAGAAA TTGTCTGGTA TGACGGGTAC	3900
AGGTAAGACT GAGGAAGAAG AATTCCGTGA AATCTACAAC ATTCTGTGTA TTCCAATCCC	3960
AACAAACCGT CCTGTTCAAC GTATTGACCA CTCAGACCTT CTTTATGCAA GTATCGAATC	4020
TAAGTTTAAA GCGGTGTCG AAGACGTAA GGCTCGTTAC CAAAAGGGTC AACCTGTCTT	4080
GGTTGGTACA GTAGCGGTTG AAAGTAGTGA CTACATTTCT AAGAAATTGG TTGCAGCTGG	4140
TGTTCCCTCAC GAAGTCTTGA ATGCCAAAAA CCACTATAGA GAAGCCCAA TCATCATGAA	4200
TGCTGGTCAA CGTGGTGCCG TTACCATCGC AACCAACATG GCGGGTCGTG GTACCGACAT	4260
CAAGCTTGGT GAAGGTGTTT GTGAACCTGG AGGACTTTGT GTTATTGGTA CAGAACGTCA	4320
TGAAAGTCGT CGTATCGATA ACCAGCTTCG TGGACGTTCA GGTCGTCAAG GAGATCCAGG	4380
TGAGTCACAA TTCTACCTAT CTCTGAAGA TGATTTGATG AAACGTTTTG GTTCTGAACG	4440
CTTGAAGGGA ATCTTTGAAC GCTTGAACAT GTCTGAAGAG GCCATTGAGT CTCGCATGTT	4500

959

GACGCGTCAG GTTGAAGCAG CTCAGAAACG TGTCGAAGGA AATAACTACG ATACCCGTAA	4560
ACAAGTCCTT CAATACGATG ATGTCATGCG TGAACAACGT GAGATTATCT ATGCTCAACG	4620
TTACGATGTC ATCACTGCAG ATCGTGACTT GGCACCTGAA ATTCAGTCTA TGATCAAACG	4680
CACGATTGAA CGTGTGTTG ATGGTCATGC GCGTGCCAAA CAAGATGAAA AACTAGAGGC	4740
AATTTTGAAC TTTGCTAAGT ACAACTTGCT TCCTGAAGAT TCTATTACGA TGGAAGACTT	4800
GTCAGGCTTG TCTGATAAGG CCATCAAGGA AGAGCTTTTC CAACGTTTCCT TGAAGGTTTA	4860
CGATAGTCAG GTTTCAAAAC TACGCGATGA AGAAGCAGTT AAAGAATTCC AAAAAGTTTT	4920
GATTCTACGA GTGGTGGATA ACAAGTGGAC AGATCATATC GATGCCCTTG ATCAATTGCG	4980
TAACGCGGTT GGACTTCGTG GCTATGCTCA GAACAACCCT GTTGTTGAGT ATCAGGCAGA	5040
AGGTTTCCGT ATGTTTAATG ATATGATTGG TTCGATTGAG TTTGATGTGA CACGCTTGAT	5100
GATGAAAGCA CAAATTCATG AACAAGAAAG ACCACAGGCA GAACGTCATA TCAGTACAAC	5160
AGCGACTCGC AATATCGCTG CTCACCAAGC AAGTATGCCA GAAGATTTGG ATTTGAGCCA	5220
GATTGGACGC AATGAACTTT GCCCATGTGG TTCTGGTAAG AAATTTAAAA ACTGTCACGG	5280
TAAAAGACAA TAAATGAGA TAGTTTAGAG GCGGATATCT TGTGAAAAGT AAATTTTAC	5340
TGGGTATCCG TTTGCTTTAT AAGGAGATGA GTTATGGTAT TTACAGCAAA AAGCTCTAAA	5400
ATAAATATAG AAGAAGTTCG TGCCTTGTC AATTAGAAG GTCAGGCTTT GGAGAGGAAA	5460
TCACAGCGAG ATCAAGAGCT AGAAGCCATT ATACGTGGAG AAGACCAGCG AATTCTCTTG	5520
GTAATCGGGC CATGCTCATC TGACAACGAA GAAGCTGTCC TTGAATACGC TAAGCGTTTG	5580
GCAGTCCTAC AAGAAGAAGT GGCAGATCGT ATCTTTATGG TTATGCGTGT TTATACTGCC	5640
AAACCCCGTA CCAACGGAGA TGGCTATAAG GGCTTGATTC ACCAGCCTAA CGCGACAGAA	5700
GCGCCTAGTC TTATCAATGG AATCAAAGCC GTTCGCCATC TTCACTATCG TGTCATCACA	5760
GAAACAGGGA TGACAACTGC TGATGAAATG CTTTATCCTG AAAACCTTCC GCTTGACAT	5820
GATTTGATTT CTTACATGGC AGTTGGTGCC CGTTCAGTTG AAGACCAGCA ACACCGCTTT	5880
GTGGCAAGTG GGGCAGGATT TTCTACTGGT TTTAAAAATC CAACCTCTGG AAATCTCAAT	5940
GTCATGTTTA ATGGGATTTA TGCTGCTCAA AACAAACAAA GTTTCCTTTT CTTAGGAAAA	6000
GAAGTAGAAA CAACTGGGAA CCCGCTTTCA CACGCTATTC TTCGTGGTGC TCTTAATGAG	6060
TATGGAAAAA ATATTCCCAA CTACTATTAT GACAATTTAA TTGATACCAT TGCCCAGTAT	6120
GAGAAAAATG GCTTGAAAAA TCCTTTTATC ATCATTGATA CCAATCATGA CAATTCTGGT	6180
AAGCAGTATA TTGAACAGAT CCGAATTGTC CGCCAGACCT TGATTAACCG TGCTTGGAAT	6240

960

GAAAAAATTA AGCAGTTCGT TCGTGGTTTT ATGATTGAGT CTTATCTGGA AGATGGTCGA	6300
CAAAATGAGC CAGAAGTATT TGGTAAGTCT ATCACAGACC CTTGCCTGGG TTGGGATAAC	6360
ACAGAAGCTC TTGTCAGAGA AATTTACAAA ACGTTAGGAG AATAAGATGG CATTATTGA	6420
AAAAGGTCAA GAAATCGATA TGGAAATCAT CAAGGCTGAA ACCCAATTGT CTGCGGAAGC	6480
CTTGAGACTC AAGGAAAGCC GTGACAGGGA ATTGGCAGAT ATTATTTTCAG GGGGAAGATGA	6540
CCGTATTCTC TTGGTGATTG GTCCTTGCTC TTCTGATAAT GAAGAGGCGG TCTTGAATA	6600
TGCTCGCCGT TTATCTGCCT TGCAAAAGAA GGTAGCGGAT AAGATTTTCA TGGTCATGCG	6660
CGTGATACT GCTAAGCCTC GTACCAATGG AGACGGCTAT AAAGGATTAG TTCACCAGCC	6720
AGATACTTCT AAGGCTCCAA GCCTGATTAA TGGCTTGCGG GCTGTGCGCC AGTTGCACTA	6780
CCGCGTGATT ACAGAGACTG GTTTGACAAC GGCAGATGAG ATGCTTTATC CGTCAAATCT	6840
GATCTTGGTG GATGACTTGG TCAGCTACCA TGCCGTTGGA GCTCGTTCTG TGGGAAGACCA	6900
AGAGCACCGC TTTGTGGCTT CTGGGATTGA TGCACCAGTA GGGATGAAAA ATCCAACCTC	6960
AGGAAATTG GGTGTTATGT TTAACGCCAT CTATGCTGCT CAAAACAAGC AAACCTTCCT	7020
TTATCATGGG CAGGAAGTTG AGACATCAGG TAATCCTTTG GCCCATGTTA TCCTCCGTGG	7080
AGCAGTCAAC GAGTATGGCA ATTATATGCC GAATTACTAC TATGAAAATC TACTCCAAGC	7140
CATTGAACGC TATGAAACCA TGGGACTTGA AAATCCTTTT ATCCTCATTG ACACCAACCA	7200
TGATAACTCA GGCAAGCAAT ATATGGAGCA GATTGGAATT GTTCGCCAGA CCTTGCAGAA	7260
TCGTGATTGG AATGAGAAAA TTA AAAAGAC GGTTCGAGGA TTTATGATTG AATCTTACCT	7320
AGCAGATGGT CGTCAAAACC AACCAGAGAT CTTTGGTTGC TCTATTACTG ACCCTTGCCT	7380
AGGTTGGGAA AATACAGAGG CCTTGGTAGA AGAGATTTAT GTTACCTTGA CAAAATAAGT	7440
GAAAAGGATG GAGTTGGGGA ATCTCAACTC CTTTGTATGA GAATGATAGT TGGACACGGA	7500
ATTGACATCG AAGAATTGGC TTCGATAGAA AGCGCAGTTA CACGACATGA AGGATTTGCT	7560
AAGCGTGAC TGACCGCTCA GGAAATGGAG CGCTTCACCA GTCTCAAAGG ACGCAGGCAA	7620
ATAGAATATT TAGCTGGTCG CTGGTCGGCT AAGGAGGCCT TTTCCAAGGC TATGGGAACG	7680
GGCATTAGCA AGCTCGGTTT TCAGGATTTG GAAGTCTTGA ACAATGAACG TGGGGCGCCT	7740
TATTTTAGTC AGGCACCATT TTCAGGAAAG ATTTGGCTGT CTATCAGCCA CACCGATCAG	7800
TTTGTGACAG CCAGTGTCAT TTTGGAGGAA AATCATGAAA GCTAGTCCAC ATAGACCAAC	7860
CAAGGCTCTG ATTCATCTGG GAGCTATTCG ACAAATATT CAGCAAATGG GGGCTCATAT	7920
CCCTCAAGGA ACGCTCAAGT TGGCTGTGGT TAAGGCCAAT GCTTATGGTC ATGGAGCTGT	7980
TGCCGTTGCC AAGGCAATTC AAGATGATGT TGATGGCTTT TGCGTTTCCA ATATCGATGA	8040



961

AGCCATTGAA CTCAGACAAG CTGGACTCAG CAAGCCAATC CTCATTTTAG GAGTTTCTGA	8100
AATCGAAGCT GTTGCTCTAG CTAAGAATA TGAAGGAAGT GGACCTAACT GGATTGACAG TCCACCTCAA	8160
CTGGATTCAA GCACTCTTAG ATAAGGAAGT GGACCTAACT GGATTGACAG TCCACCTCAA	8220
GATTGATTCA GGGATGGGAC GGATTGGTTT TAGAGAGGCA AGTGAGGTTG AGCAGGCTCA	8280
AGATTGCTC CAACAACACG GTGTTTGTGT TGAAGGAATC TTTACCCACT TTGCTACTGC	8340
TGATGAGGAA TCAGATGACT ATTTTAATGC CCAGTTAGAA CGGTTTAAAA CTATTTTAGC	8400
TAGTATGAAG GAAGTTCCAG AGCTGGTTCA TGCTAGCAAT TCTGCAACGA CTCTTTGGCA	8460
TGTAGAGACT ATTTTCAATG CGGTTCTGAT GGGAGATGCC ATGTATGGCC TCAATCCAAG	8520
TGGAGCGGTC TTGGATTTGC CTTATGATTT GATACCGGCC TTGACCTTGG AGTCTGCTCT	8580
GGTTCATGTC AAGACAGTTC CAGCTGGAGC TTGCATGGGC TATGGAGCAA CTTATCAAGC	8640
GGATAGCGAG CAAGTCATCG CGACCGTCCC AATCGGGTAT GCAGATGGAT GGACAAGAGA	8700
CATGCAAAAT TTCTCTGTCT TGGTAGATGG CCAAGCTTGC CCAATTGTCT GCAGGGTTTC	8760
GATGGACCAA ATCACTATTC GATTGCCTAA GCTTTATCCG CTAGGAACCA AGGTAACCTT	8820
GATTGGCTCC AATGGGGATA AGGAAATCAC TGCAACTCAG GTAGCGACCT ACCCGCTAAC	8880
CATTAECTAT GAGGTGGTTT GCCTCCTCAG CGACCGTATT CCGAGAGAAT ATTATTAGAA	8940
AAGAAAGGAG TCGAGCATGA ATCTACATCA ACCCTTGCAT GTCTTGCCTG GTGTGGGACC	9000
AAAGTCAGCA GAAAAATACG CCAAACTAGG AATTGAAAAC TTGCAAGATC TCTTGCTCTA	9060
CTTCTCTTTC CGTTATGAAG ACTTCAAAAC CAAGCAGGTG CTGGAGCTGG AAGACGGTGA	9120
GAAGGCAGTT CTTTCTGGTC AGGTAGTGAC TCCTGCTAGT GTCCAGTATT ATGGTTTCAA	9180
GCGCAATCGC CTGCGTTTTC GTCTCAAGCA GGGAGAGGTC GTTTTTCGGG TGAATTTCTT	9240
TAACCAGCCC TATCTGGCTG ATAAATAGA GTTGGGAGCA ACCCTTGCTG TCTTTGGAAA	9300
ATGGGACCGC GCTAAGGCTA GTCTGACTGG GATGAAGGTT CTGGCTCAGG TAGAAGATGA	9360
CCTCCAGCCT GTCTATCGTC TGGCTCAGG AATCAGTCAG GCCAGTCTGG TCAAGGTCAT	9420
CAAGACGGCT TTTGATCAGG GACTGGACCT CTTGATAGAA GAAAATCTGC CCCAGTCTTT	9480
ACTAGACAAA TACAACTCA TGTCCCGTTG TCAGGCAGTC CGTGCTATGC ATTTTCCAAA	9540
GTATTTGGCA GAATACAAGC AGGCTCTTCG CCGTATAAAG TTTGAGGAAC TCTTTTATTT	9600
CCAAATGCAG CTGCAGATGC TCAAGTCTGA AAATAGAGTT CAGGGAAGTG GTCTGGTTCT	9660
GAATTGGTCT CAGGAAAAAG TGACAGCAGT TAAAGTAAGT CTTCTTTTTC CCCTGACCCA	9720
AGCTCAGGAA AAGAGTTTGC AGGAAATTTT AACTGATATG AAGTCCGACC ACCACATGAA	9780

962

TCGTCTCCTA CAAGGGGATG TGGGGAGTGG AAAAACGGTA GTCGCTGGCT TGGCCATGTT	9840
TGCGGCAGTG ACAGCAGGTT ATCAGGCTGC CCTAATGGTA CCAACAGAAA TCCTCGCAGA	9900
GCAACACTTT GAGAGTTTAC AGAACCTTTT TCCCAATTG AACTGGCTC TCTTGACAGG	9960
TTCCTTGAAA GCTGCAGAAA AGAGAGAAGT CTTGGAGACC ATTGCCAAGG GTGAGGCTGA	10020
TTTGATTATA GGAACCTCAG CTCTGATACA AGATGGGGTG GAGTATGCTC GTCTTGGTTT	10080
GATTATTATC GATGAGCAGC ACCGTTTTGG TGTAGGGCAA AGGCGTATTT TACGGGAAAA	10140
AGGTGACAAT CCAGATGTCC TCATGATGAC GCGGACTCCC ATTCCACGGA CGCTTGCCAT	10200
CACAGCCTTT GGAGATATGG ATGTTTCCAT TATCGACCAG ATGCCAGCAG GTCGGAAGCC	10260
TATTGTGACG CGCTGGATCA AACATGAGCA ACTACCTCAG GTCTTGACTT GGTTAGAGGG	10320
GGAAATTCAA AAAGGTTCCC AAGTCTATGT CATCTCTCCT TTGATTGAAG AATCAGAAGC	10380
TCTAGATTTG AAAAATGCCA TTGCCTTATC AGAGGAGTTG ACGACTCATT TTGCAGGCAA	10440
GGCAGAGGTG GCTCTTCTAC ATGGTAGGAT GAAGAGTGAC GAAAAAGACC AGATCATGCA	10500
GGATTTCAAG GAGAGAAAGA CGGATATTCT GGTTCGACG ACGGTTATTG AGGTTGGGGT	10560
CAACGTCCCC AATGCGACTG TCATGATTAT CATGGATGCC GATCGCTTCG GTCTCAGTCA	10620
ACTTCACCAG CTTAGAGGTC GTGTCGGTCG GGGGGACAAG CAGTCCTACG CTGTTCTCGT	10680
TGCTAATCCC AAGACGGATT CTGGGAAAGA CCGCATGCGT ATCATGACAG AAACGACCAA	10740
TGGATTTGTC CTTGCGGAGG AAGATTTGAA AATGCGTGGT TCTGGTGAGA TTTTGGAAAC	10800
CAGACAGTCA GGACTTCCAG AGTTCCAAGT GGCTGATATT ATCGAAGATT TTCCGATTTT	10860
AGAAGAAGCA AGAAAGGTTG CTAGCTACAT TAGTTCTATA GAAGCTTGGC AAGAAGATCC	10920
AGAGTGGCGC ATGATTGCCC TTCATCTGGA AAAGAAAGAA CATCTGGATT AAGCTTTCTC	10980
TAAGGAAAAC TTATACTCAA TGAAAATCAA AGAGCAAACCT AGGAAGCTAA CCGCAGGTTG	11040
CTCAAAACAC TGTTTTGAGG TTGTGGATGA AACTGACGAA GTCAGCTCAA AACACCGTTT	11100
TGAGGTGGCA GATAGAACTG ACGAAGTCAG TAACATATAT ATACGGTAAG GCGACGCTGA	11160
CGTGGTTTGA AGAGATTTTC GAAGAGTATT AAGCTAGTTT TTAGGTTTGG CTCTTATACT	11220
AGAGTCATCA AAAAGAAACG AGGACTCTCA TATGACAGTA ACGATTAAAG TAAATTACCA	11280
AACCACTTTC CAAAAGAAGG AAGCAAAAAA CTAGTATAAA CAGAAGAGAG AGCGAAATGC	11340
TCTTTTTCG TTTCTAAAAC TACTTTCAGC CCATCATCCT AAAAGTAAAG AATCTAAATT	11400
CACTTCTAT TTACCCTTCT TTCTTGCAAT GATTACATAG ATATGCTACA GTTGTGGTAA	11460
CGATTACAAA ATAAAAGGAG CATGCTATGA AAAATCCAGC TTGCTAGAA GAAATTAAGA	11520
CCTATAGAGG AAGGGATGAG GTTCCGAAG ACTTTGATGA TTTCTGGGAT GGGGAAGTGA	11580

963

AAAATGTTTC CACGCTTCCA TCCTACCACT TGGAGGAAAG AGATTTCAC ATTCCTCAAG	11640
TCAAGTGCTA TGAGTTAACA TTTGAAGGAA GCAAGGAAGG AAAGGTCTAT GCACGCATTG	11700
TTCTTCCAAA GAGTGAGGAG AAGGTCCCAT TAATCTTCCA TTTTCATGGT TATATGGGAC	11760
GTGGCTGGGA CTGGGCCGAC ATGCTGGGCT TCACCGTAGC TGGTTACGGT GTTGTMTCCA	11820
TGGATGTGCG GGGCCAGTCA GGTTACTCAC AAGACGGCTT GCGTTCTCCT TTAGGAAATA	11880
CCGTGAAGGG GCATATTATC CGTGGTGCTG TGAAGGTCG GGACCACCTC TTTTATAAGG	11940
ATGTTTATCT GGATATTAC CAGTTGGTCG AAATTGTTGC TAGTCTGTCT CAGGTTGATG	12000
AGAAGCGTCT TTCTAGCTAT GGTGCCTCAC AAGGAGGGGC TCTAGCTCTA GTTGCAGCAG	12060
CGCTCAATCC TCGAATTCAG AAAACAGTTG CCATTATCC CTTCTGTCA GACTTCAGAC	12120
GGGTGATTGA GATTGTAAT ACTAGCGAGG CTTACGACGA ACTTTTCCGT TATTTCAAGT	12180
TTCACGACCC CTTCCATGAA ACAGAGGAGG AAATCATGGC GACCCTTGCC TATATCGATG	12240
TCAAAAATCT TGCCCATCGT ATCCAAGGTG AGGTTAAGAT GATTACGGGC TTGGACGACG	12300
ATGTTTGCTA TCCCATTACC CAGTTTGCGA TTTATAATCG TCTGACCTGC GATAAAACCT	12360
ATCGCATCAT GCCTGAGTAT GCTCACGAAG CCATGAATGT ATTTGTCAAT GACCAAGTCT	12420
ACAACTGGCT CTGTGGAAGT GAGATTCCTT TTAAATATCT AAAATAAGGA GTCGACTCTA	12480
AGCACAAAAT CTTAAAAATT ACAAACACGC ATAGTATCAG GGGATTAAGA AAACCTTTATA	12540
CTATGCGTTT TATCATGGAA ATATAGTAAA ATGAAATAAG AACAGGACAA ATCGATCAGG	12600
ACAGTCAAAT CGATTCTAA CAATGTTTTA GAAACAAATG TGTACTATTC TAGTGTCAAT	12660
CTATTATATT TATAGAATTT TTTGTTGCTA GATTGTGCAA ATTGCTTAAA ATAATTTTTT	12720
TCAGAAAGCA AAAGCCGATA CCTATCGAGT AGGGTAGTTC TTGCTATCGT CAGGCTTGTC	12780
TGTAGGTGTT AATACTTTTC AAAAATCTCT TCAAACCACG TCAGCTTCGC CTTGC	12835

## (2) INFORMATION FOR SEQ ID NO: 142:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5020 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

GGGGATATGA AGAACAAAAG AATATTTAAA GACTTCCAAG CTTCAAAAAT GAGTTTAAAC	60
ATTTACACAA GCCCCTTGTT AGCCTTTGTT TTTGTCTTCA TAGGAGAGTT TGTGGCTTTT	120

964

ACTTTGTATG GTATTGGCTT GTTAGCTCTC ATCGGACTTG CTAGAAATTT TGGAGAGGCT	180
GGTCAAAATC TTGCAAGCTA CTTGCAGACC TTGCATCAGA GCTTGACGGA TAAACAAGT	240
GACTTTCGTT TAATTTTAGG ATTACTGGCC TTTGGTTATT CTTAACACTG TGTTCAAGTG	300
GACAAGAAAA GTTGAGAAAA GACCTATTCG AACCTTGGGA TTTTATAGAG AGAATTCCT	360
CAGCAATCTT CTGAAAGGAT TTAGTCTAGG CCTGGCACTT TTTCTTCTGA CCTTGTAGG	420
TTTAGTGGTC TTAGGTCAAT ATCGTTTGA ATCCATTAC TTGAATCCTT ATTCTCTTG	480
CTTTGTGCTC TTTACTATCC CATTTTGGAT TTTACAGGGG ACAGCAGAAG AAGTGGTGGC	540
CCGTGCTTGG CTACTTCCTC AATTGGCCTC AAGAACCAAT CTAAGACTAG CTATTCTTAT	600
ATCTAGCCTG TTTTATACC TGCTTCATAT GGGCAATTCT GGTCTCAGCC CTCTATCTCT	660
AGTAAATCTC TTTTATTTCG GAGTTGCCAT GGCTCTTAC CTTCTCAAAA CTGATACAGT	720
TTGGGGTGTT GCAGGTATTC ATGGTGCTTG GAATTTTGCT CAGGGTAATC TCTTTGGGAT	780
TTTAGTTAGT GGTCAACCGT CAGAACGTCT CTGATGACCT TTTTACCACA AGGCAATCAA	840
GATTGGCTAT CAGGTGGTTC TTTTGGCATA GAAGGTCCA TTATGACAAG TCTGGTATTA	900
CTACTGCTGA TTGTCTATCT TGCTAATAAA TTAAAGAAAG AAAATGAAAG GATGTGACTT	960
CGGTCCGTCC TTTTCTTCGT GAAATACTA TAAGTATGCT AAAATAGGAA TAGCACATGG	1020
AGAGAGGATT CTTATGATCA ATCATTAC AGATAATCAA TTTAACTAG TATCAAAATA	1080
TCAACCATCA GGAGATCAAC CCCAAGCTAT CGAGCAGTTG GTGGATAACA TTGAGGGGGG	1140
AGAAAAAGCT CAGATTCTGA TGGGGGCGAC TGGACAGGG AAGACCTATA CTATGAGTCA	1200
GGTCATTTCT AAAGTCAATA AACCAACTCT GGTATTGCC CACAATAAAA CTCTGGCTGG	1260
TCAGCTCTAT GGGGAGTTTA AGGAATTTT CCCTGAAAT GCAGTTGAGT ATTTCTGATC	1320
CTACTATGAT TATTACCAGC CAGAGGCTA TGTCCCTTCT AGCGATACCT ATATTGAGAA	1380
GGATAGTTCT GTCAATGACG AGATTGACAA GCTTCGCCAC TCAGCTACCT CAGCCCTTTT	1440
GGAGCGTAAT GATGTTATTG TCGTGGCCTC AGTCTCTGT ATCTATGGTT TGGGTCGCC	1500
CAAGGAATAC GCTGATAGTG TCGTTAGTCT CCGTCCTGGT CTAGAGATTT CTCGTGATAA	1560
ACTCTGAAT GACTTGGTCG ATATTCAGTT TGAACGTAAT GATATTGATT TCCAACGCGG	1620
AAGATTTTCG GTTCGTGGG ATGTGGTAGA GATTTTCCCA GCTTCCCGAG ATGAACATGC	1680
CTTTCGAGTA GAATTTTTTG GAGACGAAAT TGACCGTATT CGTGAAGTTG AGGCTCTGAC	1740
AGGTGAGGTG TTGGGAGAAG TGGATCATTT AGCGATTTTC CCAGCGACAC ACTTTGTGAC	1800
CAATGACGAC CACATGGAAG TTGCCATTGC AAAGATTCAG GCCGAGTTGG AAGAACAATT	1860
AGCTGTCTTT GAAAAGGAAG GTAACTGCT TGAAGCCAG CGTTTGAAAC AGCGGACAGA	1920

965

GTATGATATC GAAATGTTGC GTGAGATGGG CTATACCAAT GGGGTTGAAA ATTATTCTCG	1980
CCACATGGAT GGACGGAGCG AAGGAGAGCC TCCTTATACG CTTCTCGACT TCTTCCCAGA	2040
TGATTTCTTG ATTATGATTG ACGAGAGTCA TATGACCATA GGGCAAATCA AGGGCATGTA	2100
CAATGGAGAC CGTTCGCGTA AAGAAATGCT GGTAAATTAT GGTTCCTGTT TGCCGTCTGC	2160
TTTGGACAAT CGTCCTCTCC GTCGGGAGGA GTTTGAGAGT CACGTTTCATC AGATTGTTTA	2220
CGTTTCAGCG ACACCTGGTG ACTATGAAAA TGAACAGACC GAGACAGTGA TTGAGCAAAT	2280
CATTCGTCCA ACGGGACTCT TGGATCCAGA GGTGGAAGTC CGTCCGACTA TGGGACAGAT	2340
TGATGACCTC TTGGGTGAAA TCAATGCCCC CGTTGAAAAA AATGAGCGTA CCTTTATCAC	2400
AACTTTGACC AAGAAAATGG CAGAGGATTT GACCGACTAC TTCAAGGAAA TGGGTATCAA	2460
GGTCAAGTAC ATGCACTCGG ATATCAAGAC CTTGGAACGG ACGGAGATTA TCCGTGACCT	2520
GCGCTTGGGT GTCTTTGATG TCTTGGTCGG AATTAACCTG CTCCGTGAAG GAATTGACGT	2580
TCCTGAAGTG AGCCTCGTAG CTATTCTCGA TGCTGACAAG GAAGGTTTCC TTCGCAACGA	2640
ACGTGGACTC ATCCAGACCA TTGGACGTGC TGCACGTAAT AGCGAAGGTC ATGTTATCAT	2700
GTATGCGGAC ACGGTTACCC AGTCTATGCA ACGTGCTATC GATGAACTG CCCGCCGTCG	2760
CAAAATCCAG ATGGCCTATA ATGAAGAACA TGGTATCGTT CCACAAACCA TCAAGAAAGA	2820
AATCCGTGAC TTGATTGCTG TGACCAAGGC AGTTGCTAAG GAAGAAGACA AGGAAGTCGA	2880
TATCAATAGC CTCAACAAAC AAGAGCGCAA AGAACTAGTC AAAAAGCTTG AGAAACAAAT	2940
GCAAGAAGCA GTTGAAGTGC TTGACTTTGA ACTAGCAGCT CAGATTCTGT ATATGATGCT	3000
GGAAGTCAAG GCCTTGGATT AGGGGAATAG TATGATTTAT TTAAGAAAGT TAAAGAAAGA	3060
AGATTTGATG TCTTTATGGG AAATGGCTTA TTCACAGCTT AATCCAGTTT GGAAACAGTA	3120
TGATGCTCCC TATTATGATG ATTATCAGTA TTTTCAAAT TTTAAAGAAT TCGAACTACA	3180
AAAATCAGAA TCCATTTTAA GCAACTCAA TCGCCTTGGT ATTTTGTGTG ATGATAAACT	3240
AGTTGGGACT GTTTCGCGTT ATTGGGTATG TAAAGAAACA AGATGGATGG AATTGGGAAT	3300
TGGTATTTAT GATAAAAAAT TCTGGAACAC TGGTATTGGG AAAGTTGCTA TGTTGCAGTG	3360
GATAGATAGG ACGTTTCAGG ATTACTTGGA GTTGGAGCAT CTGGGTTTGA CAACTTGGTC	3420
AGGAAATATT GGTATGATGA AACTTGCTGA AAAATTAAGA ATGAAAAAAG AAGCTCATAT	3480
TCCAAAAGTT CGTTATTATC AAGGTAAATA TTTTGACAGT ATTAAATATG GTATTTTGAG	3540
AGAAGACTGG GAGAAAAATA ATGACGGTTA TTATCAAATC AATGGAAACT CCTGAAGAGA	3600
TAGAAGGTAA ATCCTTCGTT CACTGGCAAA CGTGGAGAGA GGCTTATGAT GACCTTTTGC	3660

966

CTGCGGAATT TCAGGAGACA ATGACATTAG AAAGATGTCG ACTCTTTAGT CAAAAGTATC	3720
CAGAAAATAC ATTGATTGCG ATGGATGGTG TGAAGATAGT TGGTTTTATA AGTTATGGCA	3780
ACTGTCGTGA TGAGACTATT CAAGCTGGTG AAATTATTGC TTTATATGTT TTAAGAACT	3840
ATTATGGAAA AGGAATCGCA CAAAAGTTAG TGAAAGCAGC TTTGACTGAT CTTAATCATT	3900
TTTCTGAAAT TTTCTTATGG GTATTGAAAG ATAACAAGCG CGCCATTGCT TTCTATCAAA	3960
AAATGGGTTT TACTTTTGAT GGACAAGAAA AAATACTTGA ACTTGGAAAG CCTATAAAGG	4020
AAAAACGGAT GGTATTCTAT TCTAAATAAT TCTCAAAAGT AAAAGCTAAT ATGGTACCAA	4080
GTCTGAAAAT TTAATAAATT AGAAAGCGAG TAAATTTATG TCCCGTTCCC AATTAACAAT	4140
TTTAACAAAT ATCTGTCTGA TTGAAGACCT CGAAACTCAG CGCGTGGTGA TGCAGTATCG	4200
CGCCCCTGAA AACAAATCGCT GGTCTGGTTA TGCCTTTCCT GGAGGTCATG TAGAAAATGA	4260
TGAGGCTTTT GCGGAGTCTG TCATTCGTGA AATCTACGAA GAAACAGGGT TGACTATCCA	4320
AAATCCTCAA CTTGTCGGCA TTAATAATTG GCCACTAGAT ACAGGTGGGC GCTATATTGT	4380
CATTTGTTAT AAGGCGACTG AGTTCTCTGG TACCCTTCAA TCTTCAGAAG AGGGAGAAGT	4440
TTCTTGGGTG CAAAAGACC AGATTCCAAA CTTAAATCTG GCCTATGATA TGCTACCATT	4500
GATGGAAATG ATGGAAGCTC CCGACAAGTC AGAGTTTTTC TACCCTCGCC GTACAGAAGA	4560
CGATTGGGAA AAGAAAATCT TCTAGTCTTT TACTAAATAA CCTAGCTGAT CCAAGGCCTC	4620
CTCGATATAG TGGAGGTCTT GTTGTGTCTC GGCTTCAACT AGGTGATAAT GAATACCATC	4680
TGTTAACTCA GAAATTGGCT TAAAGTCAGA ACGTTCAACT TGTCTAGAA AATGTTGCAC	4740
GTCGCGGCGA CAGGTCAGTT TTAGTAAGGT TTCAATCTCT CCATAAACAG GATGATCAAT	4800
CAAGATATTT TGAACGCGAC CACCATTATC TACGATAGCA AGTAATTCTC GTCCAATTTT	4860
TTCAACTTCA TGCTTGACCT TAAATAATTT GTGATGATAA GTATTTGCAT TAGCATCTTT	4920
ATAGATATAA CCACGATTGG TAGATAGAAT TGGAGATCCA TCAGCTCTTA AAATGCAAT	4980
ATCTTGAACA ATAACCTGTC GAGTGACATG AAAGTGCTCA	5020

(2) INFORMATION FOR SEQ ID NO: 143:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4965 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

AAAAAGTGGC AATCCATTGA TTGGCCACTT CATTTAGAGA ATTATCGTCT CGCCCTTGAA	60
---	----

967

GAAGAAGGTC GTGTAGTACT TGAGTTACTG CTATCGCTAG AACTACTACT TGAAGTGGTG	120
GAGCTGGATG GAGTTGGTAG ACTCCCCACA ATACTAGACC AAGCATTCTG ATAATCCGCA	180
TCACTTCCGC CAATAGCAAA GCGATAACTT GTCGCTGGCG CTCCTGACTT ATTAGCCCAA	240
TAGCTGGTAA CAGTCGAACC TGTGACCTCT ACTTCTTTTC CTTCAACAGA AACCTTCTCT	300
GGTTTTTGAC CTGTTGATTT CAAGACTTCC GATTTCACTA CACTAGGATC TAAAGCAAAG	360
CGCTCGTTCC CCCAAATGCT TGGGGAAGCT TGCTGAATCG CATTTACCAG ATGAGCCATG	420
TAATTAGAGT TATTAGAATA ACCTGCTCTA CGTGACAATG AATGATTATC ATCATGCCCA	480
ATCCAGCCAC CTAGGGTTAA TCTAGGTGTC GAAAGCATGA GCCACATATT TTCGTCTTGG	540
TTGGTTGTAC CAGTCTTCCC AATCCAATCT GCATTAGCCA GAGTAGGATT TAAAGAAGTC	600
AGGTTAGACT TGAAGGTTGT TGTCACACGA GAGGATAGAA CTTCTCGTAG CAATCCCTGC	660
ATAATCGTCG CAGTAGCTTT TGAATAGACT TGAACCGGTT TATCCTGATA CTCATACACC	720
ACTCTACCAT CTGCTGCTTC AATCTTTGAA ATCACATGCT TCTGATGATA AACTCCATTA	780
TTAGCTAAGG TCTGATAGCC ATTGGTATGC TGGGCAACTG TGACTTCAAT ACCACCACCC	840
ATTGGCAAGC TCTCAATACC GTACTCAGGA ATCTCGTAAC CCATCTTTTC CATATAACCC	900
TTGACATCAA CACCCTTTTC ACGGAGCATA CGATAGGTCC AGTAAGCAGG GATATTCCAT	960
GAATAGTTCA GAGCTTCTCC CAAGGTCATC ATTCTGTTC CCTTGCTATT AGCATAkata	1020
ATCGGATTGC CATTAGCAAA GTTTGTTGGA TAGTTAGATA GAATCGTTTC ACTTCCCATC	1080
AAGCCCTGGT CAATAGCAAT ACCGTAGGCC AGCAAGGGCT TGGTAGTAGA AGCTGGCGAA	1140
CGTTTGGTAT CAAAGGCATG ATTATTTTGA TTTTCTTGAT AATTACGACC ACCTACAAAG	1200
CCTAGAATAG CACCTGTTTG GTTATCCATC AAGACATTCC CTACTTCTAC ACGACCTGTT	1260
CCATCGTCTA AAAGATAGCC ATAATCAGCA ACCGCACTTT GCATGGCAGA ATGAATTTTC	1320
TGATCTATGG TAGTAGTAAT CTTATAACCA CCATTTTCAA TTTCTTGGC TGCCAAATCT	1380
CGATAAACT TCTGAGTTGC CTCATTTTTC AACTCCTTAG CGGAGACATT GTCTCTCTGA	1440
GCTAGATAGT CATAcATACG TTCTTGAGCT TCTGCCAAAG TTGTAAAGTA TAAATAGTCT	1500
CGTGAAATTC CTGTAACCGT GCCCGATGGT AAAAAGTCCT GTTTAAGGTC ATAATCCTTG	1560
TACTGAGAAT ACTCGTCTTT GCTTAATGCA CCGTACGAT ACATACTGTA AAGAACTGCC	1620
TTAGCCCGTC TTAAGCCAAT TTCTAGGTCT TCATCACTCT TCAACTCCCC AGTATTTTCA	1680
TAAGGAGAGT AAGTAATGGG ACTCTGTGGA AGTCCTGCTA AAAATGCTGC TTGAGGAACA	1740
GTCAACTGAC TGGCATCTAC ACCGAAAATT CCCTCAGCTG CTTGCCGAGC CCCTGCAATA	1800

968

TTCTGTCCCT TATTATTTCG GCCAAAGGGA GCCACATTGA GATAGGTCGT TAAAATCTCA	1860
TCTTTATPCA TGGCGCGTTC CAAGGCAAGA GCATCCACAA TCTCTGCCGC CTTACGAGCC	1920
AAGGTCGGCG CATCCCCAAC CACCTGCTGT TTAATTAGTT GCTGGGTCAA GGTGAACCC	1980
CCACTAGAGG AACCCAAACC TACAAATTC CCCAAGGTCG CACGAATCAC CGCCTTGGGT	2040
ACTACACCCT TATGTTCTTT AAAGTGTTC TCTTCTGTCG CAATGATAGC CTTCTTCAGA	2100
TTTTCCGAAA TTGCTCAGA TGAGATAGAA GTGCGCAACA AATCACTCTC TATGGAAGCA	2160
ATCACCGTCC CGTCCGAATA GGTAATCTCT GAAATAGAAG AGATGTCCTT GACCTGATTC	2220
ACCAATTCTT CTGTCTGAGG CACCCGAACC TTGTCAAATA AGGCCACTCC GTATCCCAA	2280
GCAATCCCAG CTCCCAACAT TCCTCCTAGA AAACCGAGTA CAAAGAGTAA GTTAAATAAG	2340
GCTTTTATAC TCAGTAAAAT AGCTGGGAAA ATGACTGACT TATCTAAGGT TTTAGATTTT	2400
TTGGTACTTG AACCTTCTT GCCAGGTCTA GCTGATTTT TATTTTTTTG TTTTGTCTGG	2460
AAAAATTCCA GCATTTTTCG TTTTAATTCA TTTAATTGAT TTTGCATGGA TTTCTCACT	2520
TTATCTATTA TACCACAAA GGGAAATTT CAATAAAATA GCCACTTCT TCCCTATTCT	2580
GCTAGGCTAT TGCCCAAGTT TGTGATACAA TAGGTAGAAA CAATAATTTT AAAAAGGAGA	2640
AAAAACACAT GCACATTTT GATGAGCTAA AAGAGCGTGG TTTGATATTT CAAACGACTG	2700
ATGAAGAAGC TTTGCGTAAA GCCCTAGAAG AAGGTCAAGT TTCTTATTAT ACTGGCTACG	2760
ATCCAACTGC TGACAGCCTT CACCTAGGCC ACCTTGTCGC AATCTTGACA AGTCGTCGCT	2820
TGCAACTAGC AGGTCACAAA CCTTATGCGC TCGTTGGCGG TGCTACAGGT CTCATCGGAG	2880
ATCCGTCCTT CAAAGATGCT GAACGTAGTC TCCAAACAAA AGACACAGTA GATGGCTGGG	2940
TCAAGTCTAT CCAAGGACAA CTTTCTCGTT TTCTTGACTT TGAAAATGGC GAAAACAAGG	3000
CTGTCATGGT CAACAACTAC GACTGGTTTG GCAGCATCAG CTTCAATTGAC TTCCTCCGTG	3060
ATATTGGAAA ATACTTCACG GTCAACTACA TGATGAGTAA GGAATCTGTT AAAAAACGGA	3120
TCGAAACAGG AATTTCTTAC ACTGAGTTCG CTTACCAAAT CATGCAAGGG TATGACTTCT	3180
TCGTCCTTAA CCAAGACCAT AATGTCACTC TTCAAATCGG TGGTTCTGAC CAGTGGGGAA	3240
ATATGACAGC TGGTACCGAA TTGCTTCGTC GTAAGGCGGA CAAGACTGGT CACGTTATCA	3300
CTGTTCCACT AATCACAGAT GCAACTGGTA AGAAATTTGG TAAATCAGAA GGAAATGCCG	3360
TCTGGCTCAA TCCCGAAAAG ACTTCTCCAT ACGAAATGTA CCAATTCTGG ATGAACGTGA	3420
TGGACGCTGA CGCTGTTGCG TTCTTGAAAA TCTTTACTTT CTTGTCACTT GATGAGATTG	3480
AAGATATTCG TAAACAATTT GAAGCAGCGC CACACGAACG CTTGGCTCAA AAAGTCTTGG	3540
CTCGTGAAGT TGTTACACTT GTTCACGGAG AAGAAGCCTA CAAAGAAGCA CTTAACATCA	3600



969

CTGAGCAACT CTTTGCAGGA AACATCAAAA ACCTTTCTGT CAAAGAGCTC AAACAAGGAC	3660
TTCGTGGTGT GCCCAACTAC CAAGTACAGG CAGACGAAAA CAACAATATC GTGGAAGTGC	3720
TGGTCTCATC TGGTATAGTT AACTCAAAAC GCCAAGCCCG TGAAGACGTC CAAAACGGAG	3780
CCATCTACGT AAACGGCGAC CGCATCCAAG AGCTTGACTA TGTCTTGAGT GACGCTGATA	3840
AGTTAGAGAA TGAAGTACT GTTATCCGTC GTGGGAAGAA AAAATACTTT GTATTGACTT	3900
ACTAAACTAT TCAACATTTA TCTATAAACA AAGGAGTTAA CCTCGAGAAA GGTAAGTCTT	3960
TTTTGCTGTT AATAACTCTC ATCTATCTAT TTTAATAGA CAGGCTACGC AGGACAATGC	4020
GCAAGGTTGT TAGATTATGT AAGATAGAGA GATTTGAAGG ACTGAACCAA TTAAATAAGC	4080
CAAAGCCAAT CAAACTACTA TTTACGACAA CGGTATCCTG AATATTTTTC TTGATGAGTG	4140
TTTGCAAAGA TGATGATAAC GAATCCAAC CTGGAAGAA ATCCAAACGA TTATCTAACA	4200
ATAAGATATC ACTCATCTGC TTAGAAATAT CTGCACTCTC ATTCATCACC ACACCGATAT	4260
CTGATAGAGT TAAAGCCGCT GAGTCATTCA ATCCATCTCC AACCATCAAA ATAGTGTGAC	4320
CTGCTTTCTG CAGTTTCTCT ACTAACTCAA ATTTCCCATC AGGTTTCAAG TCTGTATAGA	4380
CCTGATCAAA GGGCAAATCT TTGACTAATT CCTCTGTCCT AATCAAGGTG TCTCCTGTTG	4440
CCAGAATCAA TTTTTCCTCC TGTGCCTTAA GTTTATCCAA GGCTGTTTTT GCTTCTTTTC	4500
TCAAAGGAGT ATGAATGCAG AACATTCCAA TCAATTCATT TTGATAAGCC AAGAATAAGA	4560
GATTGTAGTG ACTCTGTAC TCTTCAATTA AAGCATTTTG TTCTGAACTG ATATGAATCT	4620
GCTCATCTG CATCAAGACA TAATCCCAA TAAGAACTGG TTGGCCATCT ATATGAGATT	4680
TGATCCCTT GCTTGCAGTA TATTGGAGTT TCCCATGCAT TTCCTCATGT TCAATTCCTT	4740
CTATCTCAGC TTGCTTGACG ATGGCATTAG CAATAGGATG ATAAATGTGT TCCTCAAGAC	4800
AGGCACTGAT TCTGAGAATA TCTTCCTCAC TATAGTCTCC AAAAGGTAAC ACCTTTTCAA	4860
CTATAGGATA ACTAGTTGTG ATTGTTCTG TCTTATCAAA CAAGAAAGTA TCAACTTCCA	4920
GATATTTCTC CCTGTTGTGG CCTCTGGCTG TCATCTCTGT GCTGG	4965

(2) INFORMATION FOR SEQ ID NO: 144:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3232 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

970

CAGGGGCGTA TTACGTGACA ATTCAATGTA GGCTGTGCGCT ACTTGCGCCA AAACAAGGAT	60
TCGATAATGT CGGATGATAC TAACGATTAA ACCGAGCAGA AAGGATCCCA AAATTCCCCA	120
AACTGCAATA TGCAAGGTCA GAAAGAATGC CTTTGTGATAT AGTGGTAGAT ATTGTTC AAC	180
AATGGATCAA TCCAAAAATA GAACCTCCCA TCTAGAAATA ATACAGTTAT TGTAGCACTT	240
AAAATCTTCT TTGGATAATA TCTATTTTTT ATTGCCGTTA TAAGGATTTT TATCATAGAC	300
ATAAAATTTC TGAAATTTC AAACAAAATA TTTTAAAAGT TTTGAAAAAG AGTTAAGATA	360
TTTTTGTAAT ACACAAAGTA AACGCTTACT TATTAAGGAG GACATTTTAT GTCATACAAA	420
ACAAGCAATG CAGAAGGTCA TGTAAGTTTC ATCAATACCT ATGATTTGGA GCCAATGCGG	480
CAACAAGTTA TTCCTAAAGC AGCATTTGGC TATATCGCTA GTGGGGCGGG AGATACTTTC	540
ACTTCTTTCC AGTGATTTTA GCGTCAGGTT CTTTTAGTT TTTAAAGATT ATCCGTGAAT	600
TTCTTGCTTA TTTATGATAA AATGGGAGTG TCGCAAAAAA TGACTCATCG TATTCAATTT	660
TGAGTAAAC TAGGAGGATC CCATGTCTAC AGAACATATG GAAGAACTAA ATGACCAGCA	720
GATCGTTCCG CGTGAAAAA TGGCTGCGCT CCGCGAACAA GGAATCGATC CTTTCGGAAA	780
ACGTTTTGAA CGTACTGCAA ATTCACAAGA ATTAAGAGAT AAATATGCCA ACCTCGATAA	840
AGAACAATTA CACGATAAAA ACGAAACAGC TACTATCGCA GGACGCTTGA TAACCAAACG	900
TGGTAAAGGA AAAGTTGGTT TTGCCACCT TCAAGACCGC GAAGGCCAGA TTCAGATCTA	960
CGTTCGTAAG GATGCTGTCG GTGAAGAAAA CTACGAAATC TTCAAAAAAG CAGACCTTGG	1020
TGACTTCCTT GGTGTCGAAG GTGAAGTGAT GCGTACGGAT ATGGGAGAAC TCTCTATCAA	1080
GGCAACCCAC ATCACACACT TGTCTAAGGC TCTTCGTCCT CTTCTGAGA AATTCCATGG	1140
TTTGACAGAC GTTGAAACAA TTTACCGTAA ACGTTACCTT GACTTGATTT CTAATCGTGA	1200
AAGCTTTGAA CGCTTTGTCA CTCGTTCAAA AATCATCTCT GAAATCCGTC GTTACCTTGA	1260
CCAAAAAGGA TTCCTTGAAG TGGAAACACC TGTCTTCAT AATGAAGCCG GTGGTGCTGC	1320
TGCCCGTCCA TTTATCACCC ACCACAATGC CCAAAACATT GACATGGTGC TTCGTATCGC	1380
GA CTGAGCTT CACTTAAAC GCCTTATCGT GGGTGGTATG GAACGTGTCT ATGAAATTGG	1440
CCGTATCTTC CGTAACGAAG GAATGGACGC TACTCATAAC CCTGAGTTCA CTTCTATCGA	1500
AGTTTACCAA GCTTATGCAG ACTTCCAAGA CATCATGGAC TTGACTGAAG GCATTATCCA	1560
ACACGCTGCT AAATCAGTCA AAGGTGATGG CCCAGTCAAC TACCAAGGTA CTGAAATCAA	1620
GATTAACGAA CCATTTAAGC GTGTTTATAT GGTGGATGCT ATCAGAGAAA TTA CTGGTGT	1680
CGATTTCTGG CAAGACATGA CTTTGAAGA AGCTAAAGCT ATCGCTGCTG AGAAGAAAGT	1740
TCCAGTTGAG AAACACTACA CTGAGGTTGG TCACATCATC AATGCCTTCT TTGAAGAGTT	1800

971

TGTTGAAGAA ACTTTAATCC AACCAACCTT TGTCTATGGA CATCCAGTAG CTGTATCTCC	1860
ACTCGCTAAG AAAAATCCTG AAGACCAACG CTTTACTGAC CGTTTCGAGC TCTTTATCAT	1920
GACTAAGGAG TACGGTAATG CCTTTACTGA GTTGAACGAC CCAATCGACC AACTTAGCCG	1980
TTTTGAAGCC CAAGCTAAAG CCAAAGAACT TGGTGATGAT GAAGCGACAG GAATCGACTA	2040
TGACTACATT GAAGCTCTTG AATACGGTAT GCCACCAACA GGTGGTTTGG GAATCGGTAT	2100
CGACCGTCTC TGCATGCTCC TCACTGATAC AACAACTATC CGTGATGTAT TGCTCTTCCC	2160
AACAATGAAA TAAATTCTTA TCCTCTGGGT CTTATCAGAG GATTTTTTGA TTCAAAAAGA	2220
GACTGAATTT AAGGAGAAAA TGAAGTGTAG TATATTGAAA TTGAAATAGT ACACTTTGAT	2280
TTCTAAGACA TTGTTAGAAA TTGGTTTAAA TTCCCTAAGC AATTTGTGCA TGTTTTATTT	2340
CATTTTACGA TAGTACGCTG AAACCTTTCA AAAAGTACTA GAAATTGACT TGGATTCCCC	2400
AATTGATTTG TTCAGATTCA CTATAAATAA AAAATTAATA AGTGGGATAG GAAGTTAGCG	2460
TCAACTAGGA TAGTATCTTG CTTAAACAGT ATATATGGGA TTGATATAAG TCCATAGGTC	2520
CTATTAGAGG ATGTTCTGGT GTCTTATTCA CTGTTTTTTT ATAGTATTAG TAGATAGAAT	2580
CAGCAAATAA AAACCCAAAT CATTCATACC TCTCTCAACT AGATGTAAC TACAAAACCC	2640
CTGACCTCAT GAGCCACTTT CTTCCTCCTC ATGAGGTCAG TTTTACTTTC TGCTGTTCCA	2700
GTATCGTTTT TCCTCGCTAG ATTCCTCAA AAGGGCAGAC TCCTCCCTTG GTGCGTCACA	2760
CGATTTTTTC ATCTCGACTG TTCTTTAATG CATCATTAAC GACGCTTTTC TTCTAGGTGG	2820
TTCATAAGGA ACAGGAAGAT TCAGGTTGAC TTTTCTAATC CTAGAATAAA GTGCTGAAAA	2880
CAATTCGGAA TAGGCATAGA GACTAGACAA TTTGAGGAGC TGCTTCCGTC CTGTTCGAAC	2940
ACATTTTCCC ACCACGTGAA GAAAAAGATG GCGGAAGCGT TTGATTGTGA AAGTTTGGAA	3000
GTCACCTCCA GCTAGATGTT TGAGAAAAAG ATAGAGATTG TAGGCGATAC AGCTCATCAT	3060
CATACGAACT TCGTTTTTGA TTAAGGTTGA ACTATCCGTT TTATCGCCAA AAAATCCCTC	3120
CTTCATCTCC TTGATGAAAT TCTCGGCTTG ACCACGTCCA CGATAAAGCT GAAACTGGTC	3180
TTGGCTTGTT CCACTCGTCA TATTTGTAAC GAGAGAAATA ACATCGTAGA AC	3232

(2) INFORMATION FOR SEQ ID NO: 145:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 10711 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

972

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

CCGGAGAAAA TGATGAAAAG TTCAAACTA TTTGCCCTTG CGGGCGTGAC ATTATTGGCG	60
GCGACTACTT TAGCTGCATG CTCTGGATCA GOTTCAAGCA CTAAAGGTGA GAAGACATTC	120
TCATACATTT ATGAGACAGA CCCTGATAAC CTCAACTATT TGACAACTGC TAAGGCTGCG	180
ACACAAATAT TACCAGTAAC GTGGTTGATG GTTTGCTAGA AATGATCGC TACGGGAACT	240
TTGTGCCGTC TATGGCTGAG GATTGGTCTG TATCCAAGGA TGGATTGACT TACACTTATA	300
CTATCCGTAA GGATGCAAAA TGGTATACTT CTGAAGGTGA AGAATACGCG GCAGTCAAAG	360
CTCAAGACTT TGTAACAGGA TTAAATATG CTGCTGATAA AAAATCAGAT GCTCTTTACC	420
TTGTTCAAGA ATCAATCAAA GGGTTGGATG CCTATGTAAA AGGGGAAATC AAAGATTTCT	480
CACAAGTAGG AATTAAGGCT CTGGATGAAC AGACAGTTCA GTACACTTTG AACAAACCAG	540
AAAGCTTCTG GAATTCTAAG ACAACCATGG GTGTGCTTGC GCCAGTTAAT GAAGAGTTTT	600
TGAATTCAAA AGGAGATGAT TTTGCCAAAG CTACGGATCC AAGTAGTCTC TTGTATAACG	660
GTCCTTATTT GTTGAAATCC ATTGTGACCA AATCCTCTGT TGAATTTGCG AAAAATCCGA	720
ACTACTGGGA TAAGGACAAT GTGCATGTTG ACAAAGTTAA ATTGTCATTC TGGGATGGTC	780
AAGATACCAG CAAACCTGCA GAAAACCTTA AAGATGGTAG CCTTACAGCA GCTCGTCTCT	840
ATCCAACAAG TGCAAGTTTC GCAGAACTTG AGAAGAGTAT GAAGGACAAT ATTGTCTATA	900
CTCAACAAGA CTCTATTACG TATCTAGTTG GTACAAATAT TGACCGTCAG TCCTATAAAT	960
ACACATCTAA GACCAGCGAC GAACAAAAGG CATCGACTAA AAAGGCTCTC TTAAACAAGG	1020
ATTTCCTGCA GGCTATTGCC TTTGGATTG ACCGTACAGC CTATGCCTCT CAGTTGAATG	1080
GACAACTGG AGCAAGTAAA ATCTTGCGTA ATCTCTTTGT GCCACCAACA TTTGTTCAAG	1140
CAGATGGTAA AAACCTTGGC GATATGGTCA AAGAGAAATT GGTCACTTAT GGGGATGAAT	1200
GGAAGGATGT TAATCTTGCA GATTCTCAGG ATGGTCTTTA CAATCCAGAA AAAGCCAAGG	1260
CTGAATTTGC TAAAGCTAAA TCAGCCTTAC AAGCAGAAGG AGTCCAATTC CCAATTCATT	1320
TGGATATGCC AGTTGACCAA ACAGCAACTA CAAAAGTTCA GCGCGTCCAA TCTATGAAAC	1380
AATCCTTGA AGCAACTTTA GGAGCTGATA ATGTCATTAT TGATATTCAA CAACTACAAA	1440
AAGACGAAGT AAACAATATT ACATATTTTG CTGAAAATGC TGCTGGCGAA GACTGGGATT	1500
TATCAGATAA TGTCGGTTGG GGTCCAGACT TTGCCGATCC ATCAACCTAC CTTGATATTA	1560
TCAAACCTTC TGTAGGAGAA AGTACTAAAA CATATTTAGG GTTTGACTCA GGGGAAGATA	1620
ATGTAGCTGC TAAAAAGTA GGTCTATATG ACTACGAAAA ATTGGTTACT GAGGCTGGTG	1680
ATGAGACTAC AGATGTTGCT AAACGCTATG ATAAATACGC TGCAGCCCAA GCTTGGTTGA	1740

973

CAGATAGTGC TTTGATTATT CCAACTACAT CTCGTACAGG GCGTCCAATC TTGTCTAAGA	1800
TGGTACCATT TACAATACCA TTTGCATTGT CAGGAAATAA AGGTACAAGT GAACCAGTCT	1860
TGTATAAATA CTTGGAACCT CAAGACAAGG CAGTCACTGT AGATGAATAC CAAAAAGCTC	1920
AGGAAAAATG GATGAAAGAA AAAGAAGAGT CTAATAAAAA GGCTCAAGAA GATCTCGCAA	1980
AACATGTGAA ATAAGTGTG CAAAATATAA GAAAGGATT AGTATTTCCC TTGAATGCTG	2040
AATCCTTTTT TACATTTGTA AAGAAAGATT CTAATATGTA CGGACCCCCA AAAGTTGGAG	2100
CCTCTTTTTG TCAGAATAGA GAAAATTTTT GTTAATTTTA CTTGTTTCCT ATTGCTTTCT	2160
CAGCTATTAT TTGTTATATT AAAAGTATAA TTATTTTTTA TTTATCAGAG TTAAGCATTG	2220
CACTTTCAGA GGAAGGAGTA TTTTTTAAAA AGAAAATGTA AACGTTTGCT CAAAAATGAA	2280
AGGATTTAGA AGTTTATGAA TAAAGGATTA TTTGAAAAAC GTTGTAATA TAGTATTCGG	2340
AAATTTTCAT TAGGTGTTGC TTCTGTTATG ATTGGAGCTG CATTCTTTGG GACAAGTCCG	2400
GTCTTGCGAG ATAGCGTGCA GTCTGGTTCC ACGGCGAACT TACCAGCTGA TTTAGCTACT	2460
GCTCTTGCAA CAGCAAAGA GAATGATGGG CGTGATTTTG AAGCGCCTAA GGTGGGAGAA	2520
GACCAAGGTT CTCCAGAAGT TACAGATGGA CCTAAGACAG AAGAAGAACT ATTAGCACTT	2580
GAAAAAGAAA AACCGGCTGA AGAAAAACCA AAAGAGGATA AACCTGCAGC TGCTAAACCT	2640
GAACACCTA AGACGGTAAC CCCTGAATGG CAAACGGTAG CGAATAAAGA GCAACAGGGA	2700
ACAGTCACTA TCCGAGAAGA AAAAGGTGTC CGCTACAACC AACTATCCTC AACTGCTCAA	2760
AATGATAACG CAGGCAAACC AGCCCTGTTT GAAAAGAAGG GCTTGACCGT TGATGCCAAT	2820
GGAAATGCAA CTGTTGATTT AACCTTCAA GATGATTCTG AAAAGGGCAA ATCACGCTTT	2880
GGTGCTTTTT TGAAATTTAA AGATACCAAG AATAATGTTT TTGTCGGTTA TGACAAGGAT	2940
GGCTGGTTCT GGGAGTATAA ATCTCCAACA ACTAGCACTT GGTATAGAGG TAGTCGTGTT	3000
GCTGCTCCTG AAACAGGATC AACAAACCGT CTCTCTATCA CTCTCAAGTC AGACGGTCAG	3060
CTAAATGCCA GCAATAATGA TGTCAATCTC TTTGACACAG TGA CTCTACC AGCTGCGGTC	3120
AATGACCATC TTAATAATGA GAAGAAGATT CTTCTCAAGG CGGGCTCTTA TGACGATGAG	3180
CGAACAGTTG TTAGCGTTAA AACGGATAAC CAAGAGGGGG TAAAAACAGA GGATACCCCT	3240
GCTGAAAAAG AAACAGGTCC TGAAGTTGAT GATAGCAAGG TGA CTCTATGA CACGATTCAG	3300
TCTAAGGTCC TCAAAGCAGT GATTGACCAA GCCTTCCCTC GTGTCAAGGA ATACAGCTTG	3360
AACGGGCATA CTTTGCCAGG ACAGGTGCAA CAGTTCAACC AAGTCTTTAT CAATAACCAC	3420
CGAATCACCC CTGAAGTCAC TTATAAGAAA ATCAATGAGA CAACAGCAGA GTACTTGATG	3480

974

AAGCTTCGCG ATGATGCTCA CTTAATCAAT GCGGAAATGA CAGTACGCTT GCAAGTTGTA	3540
GACAATCAAT TGCACCTTGA TGTGACTAAG ATTGTCAACC ACAATCAAGT CACTCCAGGT	3600
CAAAAGATTG ATGACGAAAG CAACTACTT TCTTCTATTA GTTTCCTCGG CAATGCTTTA	3660
GTCTCTGTTT CTAGTAATCA AACTGGTGCT AAGTTTGATG GGGCAACCAT GTCAAACAAT	3720
ACGCATGTCA GCGGAGATGA TCATATCGAT GTAACCAATC CAATGAAGGA TTTGGCTAAG	3780
GGTTACATGT ATGGATTTGT TTCTACAGAT AAGCTTGCTG CTGGTGTTTG GAGTAACTCT	3840
CAAAACAGCT ATGGTGGTGG TTCGAATGAC TGGACTCGTT TGACAGCTTA TAAAGAAACA	3900
GTCGGAAATG CCAACTATGT AGGAATCCAC AGCTCTGAAT GGCAATGGGA AAAAGCTTAT	3960
AAGGGCATTG TTTTCCCAGA ATACACGAAG GAACTTCCAA GTGCTAAGGT TGTATCACT	4020
GAAGATGCCA ATGCAGACAA GAACGTTGAT TGGCAAGATG GTGCCATTGC TTATCGTAGC	4080
ATTATGAACA ATCCTCAAGG TTGGGAAAAA GTTAAGGATA TCACAGCTTA CCGTATCGCG	4140
ATGAACCTTG GTTCTCAAGC AAAAAACCA TTCCTTATGA CTTGGATGG TATCAAGAAA	4200
ATCAATCTCC ATACAGATGG TCTTGGGCAA GGTGTTCTCC TTAAAGGATA TGGTAGCGAA	4260
GGCCATGACT CTGGTCACTT GAACTATGCT GATATTGGTA AGCGTATCGG TGGTGTCGAA	4320
GACTTCAAGA CCCTAATTGA GAAGGCTAAG AAATATGGAG CTCATCTAGG TATCCACGTT	4380
AACGCTTCAG AAACCTATCC TGAGTCTAAA TACTTCAATG AAAAAATTCT CCGTAAGAAT	4440
CCAGATGGAA GCTATAGCTA TGGTTGGAAC TGGCTAGATC AAGGTATCAA CATTGATGCT	4500
GCCTATGACC TAGCTCATGG TCGTTTGGCA CGTTGGGAAG ATTTGAAGAA AAAACTTGGT	4560
GACGGTCTCG ACTTTATCTA TGTGGACGTT TGGGGTAATG GTCAATCAGG TGATAACGGT	4620
GCCTGGGCTA CCCACGTTCT TGCTAAAGAA ATTAACAAAC AAGGCTGGCG CTTTGGGATC	4680
GAGTGGGGCC ATGGTGGTGA GTACGACTCT ACCTTCCATC ACTGGGCAGC TGACTTGACC	4740
TACGGTGGCT ACACCAATAA AGGTATCAAC AGTGCCATCA CCCGCTTTAT CCGTAACCAC	4800
CAAAAAGATG CTTGGGTAGG GGAATACAGA AGTTATGGTG GTGCAGCCAA CTATCCACTG	4860
CTAGGTGGCT ACAGCATGAA AGACTTTGAA GGCTGGCAGG GAAGAAGTGA CTACAATGGC	4920
TATGTAACCA ACTTATTTGC CCATGACGTC ATGACTAAGT ACTTCCAACA CTTCACTGTA	4980
AGTAAATGGG AAAATGGTAC ACCGGTGACT ATGACCGATA ACGGTAGCAC CTATAAATGG	5040
ACTCCAGAAA TGCGAGTGA ATTGGTAGAT GCTGACAATA ATAAAGTAGT TGTAACCTGT	5100
AAGTCAAATG ATGTCAATAG TCCACAATAT CGCGAACGTA CAGTAACGCT CAACGGACGT	5160
GTCATCCAAG ATGTTTCAGC TTACTTGACT CCTTGGAACT GGGATGCAAA TGGTAAGAAA	5220
CTTTCTACTG ATAAGGAAAA GATGTACTAC TTCAATACGC AGGCCGGTGC AACAACTTGG	5280

975

ACCCTTCCAA GCGATTGGGC AAAGAGCAAG GTTTACCTTT ACAAGCTAAC TGACCAAGGT	5340
AAGACAGAAG AGCAAGAAGT AACTGTAAAA GATGGTAAAA TTACCCTAGA TCTTCTAGCA	5400
AATCAACCAT ACGTTCTCTA TCGTTCGAAA CAACTAATC CTGAAATGTC ATGGAGTGAA	5460
GGCATGCACA TCTATGACCA AGGATTTAAT AGCGGTACCT TGAAACATTG GACCATTTC	5520
GGCGATGCTT CTAAGGCAGA AATTGTCAAG TCTCAAGGGG CAAACGATAT GCTTCGTATT	5580
CAAGGAAACA AAGAAAAAGT TAGTCTCACT CAGAAATTAA CTGGCTTGAA ACCAAATACC	5640
AAGTATGCCG TTTATGTTGG TGTAGATAAC CGTAGTAATG CCAAGGCAAG TATCACTGTG	5700
AATACTGGTG AAAAAGAAGT GACTACTTAT ACCAATAAGT CTCTCGCGCT CAACTATGTT	5760
AAGGCCTACG CCCACAATAC ACGTCGTGAC AATGCTACAG TTGACGATAC AAGTTACTTC	5820
CAAAACATGT ACGCCTTCTT TACAACCTGGA GCGGACGTCT CAAATGTTAC TCTGACATTG	5880
AGTCGTGAAG CTGGTGATCA AGCAACTTAC TTTGATGAAA TTCGTACCTT TGAAAACAAT	5940
TCAAGCATGT ACGGAGACAA GCATGATACA GGTAAAGGCA CCTTCAAGCA AGACTTTGAA	6000
AATGTTGCTC AGGGTATCTT CCCATTTGTA GTGGGTGGTG TCGAAGGTGT TGAAGATAAC	6060
CGCACTCACT TGTCTGAAAA ACACAATCCA TATACACAAC GTGGTTGGAA TGGTAAGAAA	6120
GTCGATGATG TTATCGAAGG AAATTGGTCA CTCAAGACAA ATGGACTAGT GAGCCGTCGT	6180
AACTTGGTTT ACCAAACCAT CCCACAAAAC TTCCGTTTTG AAGCAGGTAA GACCTACCGT	6240
GTAACCTTTG AATACGAAGC AGGATCAGAC AATACCTATG CTTTGTAGT CGGTAAGGGA	6300
GAATTCCAGT CAGGTCGTCG TGGTACTCAA GCAAGCAACT TGGAAATGCA TGAATTGCCA	6360
AATACTTGGA CAGATTCTAA GAAAGCCAAG AAGGCAACCT TCCTTGTGAC AGGTGCAGAA	6420
ACAGGCGATA CTTGGGTAGG TATCTACTCA ACTGGAAATG CAAGTAATAC TCGTGGTGAT	6480
TCTGGTGGA ATGCCAACTT CCGTGGTTAT AACGACTTCA TGATGGATAA TCTTCAAATC	6540
GAAGAAATTA CCCTAACAGG TAAGATGTTG ACAGAAAATG CTCTGAAGAA CTACTTGCCA	6600
ACGGTTGCCA TGAATAACTA CACCAAAGAG TCTATGGATG CTTTGAAAGA GGCGGTCTTT	6660
AACCTCAGTC AGGCCGATGA TGATATCAGT GTGGAAGAAG CGCGTGCAGA GATTGCCAAG	6720
ATTGAAGCTT TGAAGAATGC TTTGGTTCAG AAGAAGACGG CTTTGGTAGC AGATGACTTT	6780
GCAAGTCTTA CAGCTCCTGC TCAGGCTCAA GAAGGTCTTG CAAATGCCTT TGATGGCAAT	6840
GTGTCTAGTC TATGGCATAC ATCTTGGAAT GGTGGAGATG TAGGCAAGCC TGCAACTATG	6900
GTCTTGAAAG AACCAACTGA AATCACAGGA CTTGCTATG TTCCGCGTGG ATCAGGTTCA	6960
AATGGTAACT TGCGAGATGT GAACTTGTT GTGACAGATG AGTCTGGCAA GGAGCATACC	7020

976

TTTACTGCAA CTGATTGGCC AAATAACAAC AAACCAAAAG ATATTGACTT TGGTAAGACA	7080
ATCAAGGCTA AGAAAATTGT CCTTACTGGT ACCAAGACAT ACGGAGATGG TGGAGATAAA	7140
TACCAATCTG CAGCGGAACT TATCTTTACT CGTCCACAGG TAGCAGAAAC ACCTCTTGAC	7200
TTGTCAGGCT ATGAAGCAGC TTTGGTTAAG GCTCAGAAAT TAACAGACAA AGACAATCAA	7260
GAGGAAGTAG CTAGCGTTCA GGCAAGCATG AAATATGCCA CGGATAACCA TCTCTTGACG	7320
GAAAGAATGG TCGAATACTT TGCAGATTAT CTCAACCAAT TAAAAGATTG TGCTACGAAA	7380
CCAGATGCTC CAACTGTAGA GAAACCTGAG TTTAAACTTA GATCTTTAGC TTCCGAGCAA	7440
GGTAAGACGC CAGATTATAA GCAAGAAATA GCTAGACCAG AAACACCTGA ACAAATCTTG	7500
CCAGCAACAG GTGAGAGTCA ATCTGACACA GCCCTCATCC TAGCAAGTGT TAGTCTAGCC	7560
CTATCTGCTC TCTTTGTAGT AAAAACGAAG AAAGACTAGT ATTTAGTAAA ACCTCTTAAC	7620
AAGATTACGG AAGCAGTCTC TATCTTTTCC AATGAGGTTT ATAGTACAGA AAAAGCCTGA	7680
GAAGATGTCT TCTCAGGCTT TTGTTAAGCA CATAAATACA ATAGTGCTAT GACAAAATCA	7740
CCCAGAAAAA TCTGGGTGAT AAATGTTATG GTTGTGCTGG TTGAGGATTG TGATTTTGTT	7800
GATCAGGGGT TGTATTTGAT TGTGCGTAT TATTGTTAGG ATTGCTAGTC GTACTATTAT	7860
TTGTGCTTGG AGTGGTTGAG CTAGACTGTG AAGTTGAACT ATCTGATGAT GAGCTTGAAC	7920
TTTCAGTTGA TGGGGGTTGT TGTGGAGCAG GTGAGTTCCA CGTAGAACGA GCACCATTTT	7980
TAAATACGAA TTCTCCATTT CTGTAGAGCC CCTCTGGTAT ATTCCAATCT TCTGGATTGC	8040
TTCTTCAGCA CAGGTAGGTC ATCATAGAGC GGTAAACTTT GGCAGCGACC GTAAGGCCAT	8100
TGCCTACAAG TGGTGTCAGA CGGTTAGAAT AGCCTGTCCA TACAGCCATT GAATATTTAC	8160
GCGTATAGCC AGCAAATAGT TCATCAGGTG CTACAAATTG AGAGGTCTTG ATGTGGTTTT	8220
CAATTTCTCT GTCTGTATAG TTAGAGGTTT CTGTTTTACC AGCCTGAGGG AGCCAAGCAA	8280
GATAGGCATT TCGTCCAGTT CCATAAGTCA AGACTGTTTT CATCATGTCG GTCATCATAT	8340
AGGCTGTCGT TTCCTTCATG GCACGAGTTC CGACATTAGA GAACTCTTTT TCACTCCCAT	8400
CACTAAAGAC GACTTTATGG ATATACATTG GTTTATAGTA AGTTCCACCA TTTGCAAAGG	8460
CAGCGTAAGC AGCAGCCATC TTTTCACTAC TTGCTCCATA TTTTTGTCT GATTCGGTTG	8520
TGTTACTTGA AATGGCATTG GAGTAGTGAA TACTTGGGTA GTCGATTCCT AGACCATTGA	8580
GGAAAGTCTT GCGCGGTTG AGTCCGACCT TGTTTAGAGT TTCCACGGCT GGGACGTTTC	8640
GCGATTGTTG CAGGGCGTAT TGCAAGGTGA TGTGCCCCA GTAGCCCCA TCCCAGTTAT	8700
AAACAGGAGT ATTTGTCCCA GGGTAGTTAT AGGGCTCATC GTGAACGATA GTAGCAGTTG	8760
AATCGTAGAC ACCGTACTCC AAGGCAGGAG CATAGTCTGT GATCGGTTTC ATAGTTGATC	8820



977

CCCAGTCGCG GTTTGTTTCT ACTGCTTGGT TAATTCCGAA GGAAACATTA CTTGACTGAT	8880
GGCGTGCTCC TAGCTGGGCA ATGACTTTAC CGTTAGAAAC ATCAACAATG GTAGAAGCGA	8940
CTTGCAATTC ATCGTCTGGA TAGGCAACGT ATTCTGTCTGT ATTGTAAATA TCCCACAGAT	9000
GTGTTTGAGC TTCTTGGTCT ACATTTGTGT AGACATCCAT CCCAGTTGTG AGTAGGTTAT	9060
AGCCTGTTTC TTCTTCAACT TGATTGATGA CTTCTTGAG GTAATTATCC ATGTAAGCAG	9120
GGTAATTACT TGCTGATTTG AGACTTTGTA GTCCATCAGT AATTGGTGTA TTGACTGCTT	9180
TCTCATACTG TTCAGCAGAG ATGTAGCCTT GATTTTTCAT TTCAGATAAG ACCAAGTTTC	9240
GGCGGTCTTG GGCTGCTTCT GGATGTGAAT AGGGGTCATA TTGGTTTGGT GCCTGAGGCA	9300
TTCCAGCCAG CAAGGCTAAC TGAGGTAAAC TTAAATTATT GAGGTCTTTA CCATAGTAGT	9360
TTTGAGCTGC TGTCTGCATT CCATAGTTCC CATTAGACAT GTAGACCTTA TTTATATAGT	9420
AGGTCAAGAT TTCTTGCTTG GTTGCTTTTT GTTCTAACTG AATCGCTAAC CAAGCTTCCT	9480
GAGCCTTACG AGAAATAGTC TGGTCGGAAG TCGAAGTTGA AAAGTAAGTC AACTTAATCA	9540
ACTGTTGGGT GAGAGTTGAT CCACCTTGGA GGGAAATTGCT TTGCAGATTG CGCAAGAAAG	9600
CTCCCAGGAT ACGGATGGTA TCAATCCCCC TGTGGTCGAA GAAGCGATGG TCTTCGATAG	9660
AAACGATTGC CTTAACCAAA TCTGTGGGAA TATCATTAGC TTGGGCATTG ACGCGGCGTT	9720
CAGAACCCAA GTCAGCAATG AGTTGATTTT TATTGTCTGA GATTTTACTA GAAGTTGTTG	9780
CAACTAGTTT ACTCTCGGAT AGGCTAGGAG CCTTGCTAAC GTAGTAGAAA AAAACTCCTC	9840
CGCCTAAGAC AATGGCTGCG ATAACCAAGC TTAAGAAGCT AATGCTCAGA TACTTGATTA	9900
GGCGCAGAAT CGTTGGTTTG TTCATCTTGT TTTACCACCT AATAAATGTT CTTTGATAAC	9960
ATTGAGATAA GGAATTTGAG GGAAGGCACC AGCCTTGATT TCATATCCAT ATTCTCGAAT	10020
ATATTCAAGT GGCATTGATT TTTGTCCCTT ATCTTGATGA TAGAAGCGAA TCAAATCGAA	10080
TGCCGGCAAT AAGTAGGTTT CTTGCTGAGA AGAAAAGTGA AGAAGGACAA AGCAGATTCC	10140
TTGTTGGGCA AGGACTTGTT CCATATGCTG AATCTGATGT GGATGAAAAT TTTTCATCGG	10200
AATCGCACGT TTTTGTMTTG TTTCTTGAC TTCAAAGTCG ATGTAATATC CATTATAAAC	10260
GCCAGAATAG TCCGTCGTTG AAGCTTGTCG AAAATAGGCT TCAACAATCT TGGCAGGACT	10320
TCGTTGTGGA TAGTCCACTT GTACGATTTG AATAGGAGTT GGTTCCTTAT GTATAACAGC	10380
CAAGCCCTGA GACAAATAGT AGTCGTTGGT AGCATTGATC ATCTTTTCAA AGGGTACCGA	10440
GCTCGAATTC GTAATCATGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT	10500
TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG	10560

978

CTAACTCACA TTAATTGCGT TCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG	10620
CCAGCTGCAT TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTCGCGTA TTGGGCGCTC	10680
TTCCGCTTCC TCGCTCACTG ACTCGCTGCG C	10711

(2) INFORMATION FOR SEQ ID NO: 146:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11887 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 146:

TACATTCATT CCATCGGCTA CTCCATAATA CTTAGATAAA ACCATAGCTG AAGTCGAATA	60
CGGATACTGT AAAGTATTAT CAATTTTAAT CAAATCATCA TTACCGATAA TACTTCTGAT	120
TGCTTTTGGT AGTATGAACC ATACGTTGGT GAAATCTCAG ATAATGAAGA ATCATTAGAC	180
TCTGGACCTT TTTCTAGTGT CTCACCTACC TCATATTCTT CACCCTTACT AGAAATAACA	240
CTCAAAGCAG ATACTGTGCA TAACTGGCTA GCCAATAAAG TACTCGCAAT AATTGAAATA	300
CCCAATTTT TATAAACAGT TTTCTTCATT ATTGTATCCT CCTAATGTAA TTATAGCGTA	360
CTATTCTAAA TTTCTTAATC TACTATAGAA TCAAGAAATC TACCACCTTC TTAAATACC	420
CTCCATTATC ACATAAACAG GTAAACTTTT CAATTAATGA CTGCGCTTTT CAATCAGCT	480
AGAGGTACTT GCTTGCTTCT TTGATACTAA GTTCAGCCAT TCTTTCCTTG TTTTCTCAA	540
TAAAGCATGT TACCCAAGTG GGATTCGTTT TGGAGTAGTC TCGCAGAGTC CAGCCAATGG	600
CTTTATTGAT AAAAAATTCT GTTTGGTTCA AGTTATGAAG GAGAATCTTT TCCATTAAAT	660
GAGTATTGGT CTCTCTTTT CTTAACAACT GGTGGTCAAT AGCGACACGT CTCAGCCAGA	720
TATTATCTGA TAGGCTCCAT TTTATACTCA ATGAAAATCA AAGAGCAAAC TAGGAAGCTA	780
GCCGCAGTTG CTCAAACAC TGTTTTGAGG TTGCAGATAG AGCTGACGTG GTTTGAAGAG	840
ATTTTCGAAG AGTATTAAGA TTATTTCTTC TAGTTCAGGG TGTTCATACA CCAAATCCC	900
TACTACTCGA TCTAGGATAT CTACCGTGTC CCACAAGGAT TTTGTCACGA CTAAGTCTC	960
TAGCTTAGGC AAATCGGTTT CCTTTAGATA AGACTGCATT GCTTTCAAAT AGTTAGCAGC	1020
CACATATTGG TATTTTCTAG GATCCTTTTC CCAGCAAGTG TCTGCAAAAT CCCAATCGAT	1080
AATCTTTGTT TTTTCGCTT CTGGAAAATA TTTTATAGAG TTTATTTCTT TCAGGCACCG	1140
CAATACCTAG AAAAGAAAAT TGATGGCGCA TATAGGCTTC CATGGACCTT GCTTTTTTAG	1200
AGTCTTTTGC TGCTTCTAGC TCCTCAAGTA AATCTGCTAA ACTCATCTAA AACTCCTCTT	1260

979

GCCCCACCAA ATGGTGCTGA AAGGCATAGA CAGCCGCTG GGTACGATCG CTGACTTCAA	1320
GTTTGGCAAG AATATTGGAC ACGTGGGTCT TGACCGTCTT GAGAGAGATA AAGAGGTCAT	1380
CTGCGATGCG CTGATTTTCG TAGCCCTTGG CGATGAGTTG GAGAACATCT CGCTCACGCG	1440
CAGTCAATTC TTCATGAAGT TCCATATGAT TGGGTGGTA TTCAACCTTC TTGCTAACCT	1500
CTTGCTCAAT GGCCAGCTCG CCAGCAGCTA CCTTACTGAC GGCATGAAGC AATTCATCTG	1560
CACTAGAAGT CTTGAGCATA TAGCCTTTGG CACCAGCATC TAAGACTGGC ATGATTTTTT	1620
CATTGTCCAA ATAAGAGGTC ACAATCAAAA TCTTGGCTTC AGGCCATTCT TTAAGGATTG	1680
CTAAGGTCGC GTCAATCCCA TTCATCTCAG GCATGACAAT ATCCATGACA ATGACATCTG	1740
GACGCAGTTC CAAGGCCAAG TCAATCCCTT GAGACCCGTT GGACGCCTCA CCCACAACCT	1800
CTACATCGTC TTGGAGGTCA AAGTAGCTTT TCAAGCCCAA TCGGACCATT TCATGGTCAT	1860
CTACTAGTAA AATTTTCATC TTTACTCCTT TATCATTCCT TATCTAACAG GGAATACGG	1920
ATATCAACCG CCAGCCCTTG CTTGGGAGCT GTCAAGAGTT GAACTGTTCC AGCCATATCT	1980
TCAACCCGCT CCTTGATATT TCGCAGTCCA TAACTCAAGT CGTCTAAGCT CCCTAATGG	2040
AAACCAATCC CATTGTCCAC CACCTTCAGT TGCAATTCAA CATCTGTCTG ATAGAGGTAG	2100
ACATCTAGGC AAGATGCCTG GGCATGGCGG AGGGTATTGC TAATCAACTC TTGCAGGATA	2160
CGGAAGATAT GCTCCTCGAT TTTCTTAGGC AATTTCTGCA TATTCTGCTT GAGACTAACC	2220
CTAAGATCAC TCTTGTCCTC AAGCTCTTTT AAAAGAATTT GAATCCCTTC TATCAAGCTC	2280
TTCTGCTCCA GTTCAACTGG TCGCAAATGC AAGAGCAAAA CCCGCAAATC CTTCTGGGCT	2340
GTTTCTAAAA TAGCTGTGAC ACTCTGCAAC TGGGTCTGCA TCTTTTCTCT ATCCAATTC	2400
AAAGCCTGCT GACTGATACC CGATAAAATC ATGTGGGCCG CAAACAACTC CTGACTGACT	2460
GTATCGTGCA AATCCCGAGC AATTCGCTTC CGTTCCTTCT CGATGATTTC CTCTTCCTGA	2520
GCAAGGCTCT GATTTTCAGC TTTTGAAGA GCCTCTGTCA AAAGGTTAAG TTTACCTGAT	2580
AAGGACTTGA AACTGGCATC CAAATCTGGA TCTGCAACCT GAACCACTTC TTGCCCTGCT	2640
AATAAACGCT TGAGATTAGC CTGCATTTTT CTTAGAGAAA GCTCTTCGAT CCCTCGCCAA	2700
AACAGGGCTA AGAGACAGGT CATGGACATG CTGAAAACCA ACAATAAAAA GACAAATTTT	2760
TCTGTTTTTT CGACATCGTG CAAAAAGATA GACCAGTCAA AATCAAGTAT TTCCAGCAAG	2820
CTGTGGGAGA AAAAAAGAC AAATAGGAAG GAGGTGAGAG CAATAATGAC ATAGGCTTGT	2880
TTTTTCATCC TCTAACCACC TCCACATCAC CAATCATAGT GGTCAAGAAA ATCTTGACAC	2940
TCTTGTTACT CTTGAGATAG TCTTTTGTTC CTTGATGATA GTGTTTATTG CGGAGGGCTC	3000

980			
GCTTGGGCTG GTTGAAAAAA ATCAAATCCC CATAGAGACA GTTAACGCTG AGACTGACTT	3060		
CCACATCTAC AGGTACGATG ATTTTGGTCG TTCCTACCAT CTTTCTGAGG ATAATGACAT	3120		
TGTCATGATT GGTTAAGATG ACCCTCTCCA GATGAATAGT GTCCTTGCCC ATGAAGCGAA	3180		
AGAGATTGAT ATCATCGAAT TGGCAAGTCT GGTAGCTTGA AAAATGATGA AGATTTCCAA	3240		
ACCAACGATT TTTCTCCTTC TTAACCGTCA CGACCTCTTC AAAAACCAAA TTGGTCTGCT	3300		
CTTTTTCCTG GTTCATCATC GGGTAAAGAA GAAAGAGGCT ATAGATAACC GCAACAAAAA	3360		
TAGCTAGAAT CACAAAAGGA TTGAGCATAA CGATGAAAAA GAAGAGAATG GTTGCCGCTA	3420		
CTAAAAGAAG ATTATTTCCC TCTTTACCAG TGTAGTAGCG AATCAAAAGC AAAAAGAGGA	3480		
ATAGTATCAG CAGAAAACGC GAAAAATGCT CTGATACCAT CAAAATCAGA GCTCCTGTCA	3540		
GAAGACAGGC TTCGATAAAT AAAAAGATTT TAAATTTTCT CATAGGTTCA TCCTCTCCCT	3600		
TCTATTTTAT CACAATTCAA AAAAGTCACC TCAGTCTGAG GATGGAAAAA AGGCGCTGGT	3660		
TACGCCTTTT TCATCTGATC CTTTGTCTTCT TTTAATTTTC CATAAAGAAG ATAGTCTACT	3720		
TTTTGTAGAT CTGCTATGGT GGCACAGTTA AGGGAACACA TAATCAAGCG TAGATCTGCT	3780		
TTCCAGCCTT GGACAATGCC AATCACTTCT TCAACTGTGT AGGTTTCAAC CAATTCCAGA	3840		
ACGGTTCGTG ACAATCCAC AGCCTTAGCA CCAAAAACCA AGCACTTAAT CATATCCAGC	3900		
GGATTCCGAA CCCCTCCACT AACCAAGAGT TCGACCTTAT CTTTCCATTC TTGGGCATTG	3960		
AGAAGGGCCT GCATGGTAGA CTGACCCCAT TGATTGAGGT AATCACGCTG GCCACTACGA	4020		
CGGTTTTCGA TATAGGCAAA GCTGGTGCCA CCACGACCCG ATAGGTCCAC TGTACGAACA	4080		
CCGAATTCAT AGGCTCTTTC GATTGTCTTG GCATCCATTC CAAAGCCAC TTCCTTGAGG	4140		
ACAATAGGAA CGGGAATTTG CTTGCTATAA TCTGCTAGAT GCGATTGCCA GCTTCTAAAC	4200		
TTCTTTTCTC CCTCGGGCAT GAGTAATTCC TGCATGACAT TGACATGCAC TTGCAATAGA	4260		
ACAGGATTCA TCTCTTCTAC AGTCTGAAGT CCTAACTCGA CAGGCTTGTC CAATCCAATA	4320		
TTGGTTCCAA GGAGGAGATT GGGATGACTA GACTTGACAG AAAAAGAATC ATCCGTTGGA	4380		
TTTTTGAGGG CTGCGCTATA AGAACCCGTT ACAAATAAAA TACCACAGGA TTCCGCCACC	4440		
TGAGCCAGCT TTTGATTGAT TTCTCTTCCC TTATTACTTC CACCAGTCAT GGCATTGATA	4500		
TAAAAAGGAA AGTCCCACTT TCGACCAGCA AACTCTGTCG AAAGATCGAT TTCATCCAGA	4560		
TTGTAAAGAG GCAAGGAAGA ATGAATCAGC TCCACCTCAT CAAAGCTATT ATAGGAACTT	4620		
TTCTGCTCAA GGGCATAGAG GATATGCTCG TCCTTACGAT TTGTCGTCAT GTCCTATCCT	4680		
TTCTTGATAT AAGAGCTCAA TCCCCAGATC GGCCCAACGA TTTTTAAGG TTTTGGTTGA	4740		
TTGCGCATCA AAACCTCAGG CGATGCCACA GTCACCACCA CCAGCACCAC TACTCTTGGC	4800		

AACGGTCTGC AAATCTTGAC TGGCTTCTTT CAACTGTCTA AGCAAAGGCG TGTAATATC 4860  
TGTAATCAAG CCTTCTAAAA GCTTGCTGGC TACTTCTACT TGATCGATAA TCTTTTCTGA 4920  
TTTCCCCTGT TCCAAGGCTT CTACCAGAGA AGTCACCGTT TCTTTTGAGG AAGTTAAAAA 4980  
ATTTTGATTG ATATTTTGCT TGATTTGCTG GACCATGTGA CTCGATACAG CCACTTCCTT 5040  
GGTCCATCCC ACTAAGAAAT CACATTCTAA AGTTGGTTTC ACTTGTGAAA TTGAAAAGCC 5100  
CCAATCACGC TCCAGAACTG TCGCCAAGTT TTCTTCTTCT AACCAAGCAG CCACCTTCTG 5160  
GCGATCAAAT GACTGGTAGA GAACCAAATC CTCTGCCACA ATACAGGCAA GGTGCCCCAT 5220  
GGAACCATTG TCTCCTCGCT TAAGCAAGAC AGCGCTAGTC AGCTTGAACA AGAGCTCCTG 5280  
ATCAACAGAA ACATCATACA GAGCCAGTAA AGCCTTGACA ACCAAGACAA CGACGCTGCC 5340  
ACTAGAACCT AGACCAACT TTTTCCCTTC TCGTTCCATT TTGCCACAGA TTTCTAGAGA 5400  
AAAAGGTCTT AAATTCTGAC CACGAACAGC GAGGAAGTCT CCCATCAAAG CAATCGTTTC 5460  
TTGAATCAAG CTATAGTCAG GATTAGGCCT TAAGTCCACT GCGAAATCAA ACATATCTGA 5520  
ATAGATACGG TAGCTGTCAG AAAAAGCAAT CTCAGCCCTC ATATAGATGG GAATATCCTT 5580  
TATCAAAGCT AACTGCCCTG GCTCTAAAAT AGCATATTCA CCTGCCCAAT AGAGTTTTC 5640  
GCAAGTTTTA ACAGCAATCA TCTTGACTCA AATCCTTTGT TTTTGACACA ATCAAGCGAT 5700  
AACGATGACC GAAAATTCTT GATAAATGCT CCAAGTCTTT CTCCTGACAG AAGACCTTAA 5760  
CATTGGGACC AGCATCCATG GTAAAGTAGC AGGCCTCTCC TTTCTCACGA AGCTGGCGAA 5820  
CAAAGGCCAT AGCCTCATAA GAGGCATCCG TCAGATAAGA AAAGGCTGGA CTAGCAGTCT 5880  
TTGTCGTAGC ATGCATAGCC AGGGCATTTC TCTCCGTAA TTCTCCAATC TTGGCAAAAT 5940  
CATTTTCTTT GAGATAAATC AGCATATCCT GATAGTCCTT CTCAGACTGA CGAACCCAGT 6000  
CGTCGAAAGT CGTCGAGGTT TCCACACAAA GTTTCATCCC GTCACGGCTA GAGATTGGTT 6060  
TTTTCTTGTC CTCTAGCACC AACATAATCA TAGCTAGTTT CAAGTCTGTC TCTACAGGGT 6120  
AAATTTCTCC ACTATCCTTA TCCCAGGCTC CTAGTGGTCC ATAAAACTC CGAGAAGAAG 6180  
AACCTGAGGC AAATTTGGCT TCCTGTGCCA ACTGACTTCT ATCCAATCCA AGCTTGAAAT 6240  
AAGCATTACA AGCCTTGACC AGGGCGGACA AACCACTAGA ACTTGAGGAC AGACCCGCTG 6300  
CCGTAGGCAT ATTGTTTTGA GTATCGATAC GGACAAAGCC CTCACCAGCT GGACGATAAC 6360  
GGTCAATAAT CTTACTCATC TTGGCATGCT CGACCTCATT TTGTAGCTGA CCATTGATGT 6420  
AAAATTCGTC AGCTGTTACA TTGGCTGGTA AAGGCGACAA GGTCTCTCT GTATACATAT 6480  
TTTCCAAAGT TAGAGAAATA CTGCTAGTAG CAGGCACCAT CTCTTTTCT TTTTCTTTC 6540

982

CCCAATATTT GATAATAGCA ATATTTGCGT AGGAACGTAC TGTACAGGC TCTCTATCCA	6600
TGTCTGAACA GCTCCTTTCT CTTCTAATCT TTCTGCTAGT TCTGTGCGT GTGTCAAATT	6660
GGTTACCAAG GCTATGATAC AACCTCCTAG CCCACCACCG CTCATCTTGG CACCCAGAGC	6720
ACCATGGCTA AGAGTCGTTT CAACCAAAAA GTCTGCCTCA GGGCTACTGA CTCCAATTTT	6780
TTTTAAATGT AAATGCGCTT GACTGAGGAT TTGTCCCAGT CTTTCAGCAT CTTTTTGTA	6840
AATCGCAACT TCTGCTTGCT GGGTTAATTC TCCCAAGGCA TGCAAAAACG GTAGGGCATC	6900
CTTGCCCTTA TTTTGAACCA CTTGGATGGC TTCACGAGTA TGACCATAAA CACCCGTATC	6960
GGCAATCACC AAATAGGCGG ATAAATCCAT CTCAAGTTCT GTAAATCCTA CGTTCTTGAT	7020
AAAGCGAATA GGTGGTCAC TAAGACAGGT CTTAGCATCC AAACCACTAG GATTCATATG	7080
GGCAATCATT TCAGCTCGAT TGACCAAGAT TTCTAGTACA TCATGAGGCA GATCAGCCTG	7140
ATAGTAGTCA AATACTGCAC GAATGGCCGC TATGCTGATA GCCGCTGACG AACCCATCCC	7200
CCGTTTCTCA GGGATAGCCG AGTCAATCTC ACAACGAATG CAGGCTTCTG TGATATTCAA	7260
ATACTCCAGT GAGGCATAAA CCGCCATGGA CAAGGTATCC TCCTCATAAA GGCGCCAAGG	7320
ACTCTCTGCA GGAACCTACCT TACAGGTCAC CTCCACCTCC AAAAGAGGCA GGGAAATGGC	7380
AGGATAACCG TAAACGACCG CATGTTCCCC TATTAAAATT ATCTTACTAT GTGCCTGACC	7440
GACACCAACT TTTTTGTCA TTTTTCCTT TTAGTAGACG AAAAAACGTC TTATTTTTCA	7500
TACAAGTATT AATTCTTTCC TATCTATTTT ATTATATTTT CACAAAAAA GCGATTGTTT	7560
CCATTACAA TCGCTTCTTT CATTATTGAA CCCATTGCCC ATTATAGTTG ACAGAATAGC	7620
CATCTACGGT CGTATTCCT GCCAAGGCAC CTGAGCGCTA TAAGCGTAGT ACCATCTGCC	7680
ATTGACCTGG AACCAACCTG TCGTCATAGA ACGACGAAAG AAATCCATA CCATTAAGTA	7740
AAGAGGAAAG TCGTGAGGGA GCATGCGCCA TTGACAACCT GTTTTAGTGA CGTACAAAGT	7800
CTCATTAACA AGTACTCGTT TCGGCCATTT ATAGGTGCGG TGTGTGAGA AATAGGGTTC	7860
AATCTTCGCC CATTCTTGAT CGTTTAAATC AGTATCATAT GCTTTGCGTA TCATAACTCT	7920
AGCTTAACAT TTTTTGTGA ATACAGGTTT TAAATAATCG ACCACGAAAA TTTCTTAAGT	7980
GGAAAACGCC TTATGAAGTA TGCTACGGGA AAGTTATGCA CTTAATTTGA CAATTCAAGA	8040
TGTAAAAATA TATACTATAG TAGATTGAAA CTAGAATAGT ACACCTCTAC TTCTAAAAATA	8100
TTGTTAGAAA TCGATTGAC TGTCTGATC GATTTATCCT GTTATTATCT CATTTTACTA	8160
TAATATTTGA TAAGTTATCC TAAAAGTATT ATTATGTTGT TGTGTTATAG ATTGATTGAA	8220
TCTAACTAAA GGATCCTATT CAATTACTAG AACTATCACA TACTCAAGGT CAGCTCACAG	8280
ATGAGCAACT ATTTTGGTTA CAATGTCTAC TAAATTTAAG TCAAACAAAT AATTTAGTCA	8340

983

AAATTAAGGAA CATAAATATG ATTACAAAAC AGAATGTAAT AGTGTCTCTAC	8400
AATTTTCTACT AGATAAACT GTAAATTCTG AAGGAAGGAT CACTTCTTCA ACAGAATTTG	8460
GAAATTCGT AAGTAATTTA TCATTCCAAC ACGGAATAGC TGGACTACTG TTTCTCTAA	8520
ATAAATTGTA CCCCCAGAA CTGGATTCTA AAATACTCTC TATCATCAAG AAGGCAGTGA	8580
CAATTAGAAC GACACACACA TATGAATATC AATACTCACT GCTATTTGGT GATGCAGGCT	8640
ATCTATGGTT ACTCCTACAT TTATTTTCTA TCAGTAAAAA TCAATACTAT CTACAATTAG	8700
CAAACGTCAC CGCTAAAAA TTAATAGAGA ATTATGATAC TCTAGAGGAA ATAGACTTTG	8760
CATTGGGAAA ATCTGGTGTC CTATTATCAT TAATAAATA CTATCAATTT ACCAATGACA	8820
ATACTCTTAA AATTTTCATC CACAATAGTA TAGGGGAAAT TTATCATTAT TTCCTACAAA	8880
GAGATACAGC CAAAGAAAGC ATTTTAGACT ATAGCTTTGC TCATGGATAT TGTGGAATTG	8940
CATATGCTTT ATTTGCCTAT TCTAAAGTCT TAGAACCTTC TATGTTTTAT AATGATCTCC	9000
ATACATTCCA TACTGAATTA AAAAAATTAT TAGAAAAAGT TACTTCTAAT ACTGAAAATT	9060
TAGGAAATTT ACAACTTTCT TGGTGCAAAG GAATTTCCGG AATAATCTTA TATCTTTGTA	9120
TGTACGATTG TGACGGAAAC AAAGATATTA TTAGTAAATA TCAAGAATTT GTTTTAACC	9180
ATCATCTAAA AATGATGACA GGATATTGCC ACGGAATAAC TAGCTTACTA CAAACCACTG	9240
TCTACAATCA AAACAAATTA CTGATGAAAA AAATCCAACA GGTAATTTTA GCATGTTCTG	9300
AACGAGATGA TCACGGTTTA CTGATGTTTC AAGGAGATAG TGGTAAAGCA GATTTGTTTG	9360
ACTTCGGAAT AGGAAGCATG GGGTATATTG GTGTCTATTA AATAATAAAT TCCCATTCCA	9420
TGTGCAGACA TAAGGAGAAA AGTATGAAAT TATTTTGGAC AAACAACATA TATAGACAGT	9480
TGCTGCTAAA CAGCTGTTTT TCATCATTCG GCGACAGTAT TTTCTACCTC GCCATTATCA	9540
ATTATGTGGC TCAGTACAAT TTCGCTCCGC TAGCGATTTT ACTGATTTCC ATTTAGAGA	9600
TGGTTCCCT ACTATCGCAA CTCCTTCTCG GGATTCTAGG AGATTTTCAA GAAAATAGAG	9660
TCAAACACGC ACTCTGGATT GCCAAAATCA AAATCCTGCT CTACGCTATT TTGACAGTAT	9720
TTCTCGTCTT GTCGCCCTTT TCATTAGTTT CAGTCATTAT GATTGTCATC ATCAACCTCA	9780
TCTCTGACAC CTTGAGCTAC CTGTCTGCCT ACATGATGAA CGCCCTCTAC ATCAGTGTA	9840
TTAAGGACGA CTTGATGAT GCCATGGGGT TCAGGCAGTC TCTGATGAGG GTTGTCCGTA	9900
TTGTGCGCAA TCTGGCTGGC GCATTCCTTA TCAATGTTAT AAGTATTCAA ACTATTTCCC	9960
TTATCAACAC TCTGACTTTT GTCATTGCCT TTTTGGGCT GTATGTTATT CGACATACCT	10020
TGTATGAGGT TGAAAAAGA ATTGAAATGT CACATACAGC ACTGAGTTTT AAGAAATATT	10080

984

TTCAACATCT TAAACAGTCG CTGGCTGTGC TCCTGAGGTT AAAAGATACC GTCATACTAC	10140
TGTTTCTGAC GACCAATATG ATTGCCATCT TGGATGTGTC CCCTCGGCTG ATTGCCCTCC	10200
GCTTCATCCA ACAGACACTA GCACAACCTGA GCATTGGGCA ACTCCTCGCC CTGCTCTCCA	10260
TCATCATGTC TTGTGGAGCT ATCCTTGGCA ATATGACCAG CAGTAATCTA TTTAAAAATA	10320
TCCGTTTCAC GCACCTCTTG GTTTTCTGTG AGATTTCCCT ATTGACTCTA ATAACTAGTA	10380
TCCTTTGTCA AGCCTATATC GTAATTTTCA TGACCAGTTT CATCAGTTCT ACGATTATCG	10440
GCATTCTCAG CCCTCGCCTA CAAGCAGCTG TCTTTGCCCA TATCCCCAGT GACAAGATGG	10500
GGACGGTGGG CTCTGCTCTG AGCACAGTGG ACATTCTCGC CCCGTCCCTG CTCTCCCTAT	10560
TAGCCCTATC CATAGCATCG GCGGTTTCGG TGCAGTTAGC ATTGATATTT TTGTATCTTA	10620
TTTTAATTGC TCTTATCTTT TGTCAATGGT TAGTCAAGTT CAACACTCAT AACTAACGAA	10680
AAAGCATGTG TAGATTTTAC ATGCTTTTAA TCTCCCCAAT CGTCAGGTCA AGTACAACAA	10740
AGTCACTTCT TTGATTAAGC GAGTGTCTA ATATAATTAT AAGCGCCCTG TCATTACCGA	10800
ACCCATTCGC CATTATAGTT GACAGAATAG CCATCTACGG TCGTATTCAC TGCCAAAGCA	10860
CCTGAGCTAT AAGCATAGTA CCAGTTGCCA TTGACCTGGA ACCAACCTGT CTTATGTCT	10920
CCATTACCTG CATTTAGGTA GTACCAAGTT GAACCATCTT GATACCAACC AGTTGCCATA	10980
GCTCCTGATG AACGGAGATA GTACCATTTG TTCCCAAGGT TTTGCCAACC TGTTTTCATA	11040
TCGCCATTTG GGTGGTCTAA ATAATACCAA GTGGTACCTT CCTGATACCA GCCAGTGGCC	11100
ATTGCTCCTG AGGAACGGAG GTAGTACCAC TTATTACCTA GATATTGCCA ACCTGTTTGC	11160
ATAATACCAG TTGTTGGATC TAGGTAGTAC CAAGTCGAAT CATCGTTTAT CCACCCCGCA	11220
CGTCTTTTAC CACCAAGGTA GTTTTCTCCA TTAATTTCCG TCTTAGCTAG ATAATACCAG	11280
TTAGACTGAT CATAAAGCCA ACCTGTCTCT AAAGAATGAT TTTGATTAAA GTAATAGTTC	11340
GTATAATAAC GCTTCTCTTC TTTATCTTCT GAATCTTCAC GTTTTTCCCC GTACTTTCTT	11400
CCAACACTGT CTTTAGTTTT AATCTCTAAT GTTTTCCAAC CAACAACTC TTGTAGCACT	11460
CCATTTTTAT CGAAGTAGTA CCACTCTGAC TTTGGAAAAC CTTCTAATCT GATACCATT	11520
GGGTAAGGAC CAATTGTACT ACCTTTAGAT GGAAACGGGA TATATTGCCA GCCGACAACC	11580
ATCTCTCCAG ATAGAGAATC AAAATAATAG TACTTACCAT CAATCACTCG CCAGTAGGTT	11640
TCTTTGAGGT CCCCCTTTTT GTAGTAGGTT CTTCCGTTTT CTTGGACAAA CTGCCATCCT	11700
TCAGAATCAT CTGCAAATAC TGTACTGGTC CCTAGCAAAC CAAAGAAAAA TACTGTCACT	11760
CCAACTTGCA TAGTTTTTTT CAAAATTTTC ATCTATATAC CCTCCAATAT TAAATCCACT	11820
CACCAGATGA GCGGAAATTA TAACTTTTAC CATCGATAGT TTGGCTACCT GTAACCATTC	11880



985

CTCCAGG

11887

## (2) INFORMATION FOR SEQ ID NO: 147:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11140 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

CCGGTATGTT CTGGAATACT ACCAATCTAA GCTGGCTGTG CCCTACAGTT TTACAACCTT	60
GTACGAATAC CTTAAGGAAT ATGACCGATT TTTTCAGCTGG GTTTTGGAGT CTGGTATTTT	120
AAACGCTGAT AAAATATCCG ATATTCCTTT ATCAGTTTTG GAAAATATGT CTAAGAAAGA	180
CATGGAATCC TTTATCCTTT ATCTACGTGA ACGTCCCTTG CTGAATGCTA ATACAACAAA	240
ACAAGGTGTT TCACAGACAA CTATCAATCG AACCTTATCA GCACTTTCTA GTCTTTACAA	300
GTATCTAACC GAGGAGGTTG AAAACGATCA GGGGGAACCT TATTTCTATC GTAATGTAAT	360
GAAAAAAGTT TCCACCAAGA AAAAGAAAGA AACCTTGCT GCCAGAGCTG AAAATATCAA	420
GCAAAAACTC TTTCTAGGTG ATGAAACAGA AGGTTTTCTA ACTTATATCG ATCAAGAGCA	480
CCCACAACAG CTTTCAAATC GAGCTCTCTC ATCATTCAAC AAAAATAAAG AACGAGATTT	540
AGCCATTATT GCCCTTCTCT TGGCATCTGG TGTTGCTTA TCTGAAGCTG TTAATCTAGA	600
TCTAAGAGAT CTCAATCTAA AAATGATGGT TATTGATGTT ACTCGAAAAG GTTGCAAACG	660
TGACTCAGTC AATGTCGCTG CTTTGTCTAA ACCTTATTTA GAGAATTATC TGGCCATTCTG	720
GAATCAACGC TATAAAACGG AAAAAACAGA TACAGCCCTT TTTTAACTC TCTACAGAGG	780
TGTTCCCTAAT CGTATCGATG CTTCTAGCGT TGAGAAAATG GTTGCTAAAT ACTCAGAGGA	840
TTTTAAAGTG CGTGTAAAC CCCTAAACT GCGCCATACA CTAGCAACTA GGCTCTATGA	900
TGCGACTAAA TCACAAGTTT TAGTCAGTCA CCAACTAGGA CATGCTAGCA CACAAGTCAC	960
TGACCTCTAT ACCCATATTG TTAGTGATGA ACAAAGAAT GCTCTGGATA GTTTATGATT	1020
TTACGTATTT TAAATTATGT AAATAAATAT CAAAAAAGA AGTTGGCCAA CTTCTTTTTG	1080
ATTTATCCAA CTACCGCTTC AGCGATTCTC TCACGGCTAA TACCAGCGAA GTAGCGTGTG	1140
ATATCAATGG TTTTLAGCGC CTTAAGAACA TCTTCGCGTT CGTATTTTAC CCCACGAAGG	1200
ACATCTTCTA CTGCAGCAAC GTCTTCAATA CCAAAGAAGT CACCATAAAT CTTGATGTCT	1260
TGGATTTTTG ATTCAGTAAC GTTAGCAAAG ACTTCAACCT TACCACTAGT GAATTTGATT	1320

986

CCACGACGGA CGTTAAATTC AGGTGATTTA CCATAGTTCC AGTCCCAAGT TCCAACTTA	1380
GTATCCTTGA TGCGATTGAT TTCGGCCAAT TCTTCTTCTG AAAAGACGTA TTCAGTCATC	1440
TCTGGGTACT CTTTTTTCAT GTATTCCAAG AGTAAATCAC GGAATTTTTC GACTGTGATT	1500
TTTTTTGGTA ATTCATTGAT AATATTGGTT ACACGGGCAC GGACGGATTT CACACCTTTT	1560
GATTCAAATT TATCTTTTGA AACCTTAAGG GCATTTCGGA GGAAGTACAA ATCAACGTCA	1620
AAGAGCAAGC AACCGTGGTG CATGATACGG CCGTTGATAT AGGCTTGGGC ATTGCCACAG	1680
AACTTCTTAC CATCAATCTC AAGGTCATTA CGACCTGTGA ACTCAGCTTT AACCCCAAGT	1740
TGAGCCAGGG TATTGATAAC CGGAGTTGAG AAGCTCTTGA AGTCAAATGC CTTATTTTCA	1800
TCTTCTTTGG AGATGATCGT GTAGTTGAGG TTATTTAAAT CGTGGTAAAC AGCTCCACCA	1860
CCACTAATAC GGCGAACTAC CTCAATACCA TTTTCGCGAA CATAATCACG GTTGATTTCT	1920
TCGATAGTGT TCTGGTGACG ACCAACAATG ATAGATGGCT TGTTAATCCA AAGTAGGAAG	1980
ATTTGATCCT CATCCAAAAG GTGTTTAAAG GCGTATTCTT CCAAGGCAAT ATTAAAGCA	2040
GTGTCAATTG AATGATTGAT AATGTATTTC ATGATATCCC TTTACTTTAT ATGATAGAAA	2100
CTGGAAATAA CCTTCCAGTC TAATCTATCT TCGTTTTATT TTTTCTTAGG TGAATGGATG	2160
GCCATTCTTA GAACATCTGC AAACGCTTCG TACATCACTT CAGAGTAAGT TGGGTGCCCC	2220
TGGATGGTCT TCAGCATTTT CTCAACAGTG ATTTCCATTT CGATGATGCT TGATGCTTCG	2280
TTTATTAATT CTGCGGCTGC AGGACCAATA ATGTGTACAC CAAGGATTTT TCCGTATTTT	2340
TTATCAGCGA TAACTTTTAC GAAACCTTGA GCTGCGTCAG ATGCAATAGC ACGACCGTTA	2400
GCAGCAAAGT TAAACTTACC GATGGCAACA TCGTATTCTT CACGGGCTTG TTCTTCTGTC	2460
AAACCTACTG CTGCTACTTC AGGGAGAGTG TAGATGGCTG CAGGAGTCAA ATTCAATTTG	2520
GCAACTGCAT GATTTCTTTT AAGGGCATTT TCAGCGGAAA CTTCACCCAT GCGGAAAGCT	2580
GCGTGAGCCA ACATCTTAGT ACCGTTGATG TCACCTGGTG CATAAATGCC TGGAACTGAA	2640
GTTTCCATGT ATTCGTTGAC CTTGATACAA CCACGATCCA ATTCAAATCT AACCTCTC	2700
ATACCTTCAA GGTCTGGCAT ACGACCAATT GAAAGAAGAG CTTTGCTTGC GATGATATCG	2760
TCTTTTCCTT CAACCTTGAT ACGAAGTTGA CCATTTTCCT CAATGATTTT TTGCAGTTTA	2820
GTACCAGTCA AGATGGTCAT TCCTTTACGC TCAAGAATCA AGCGAAGGTT CTTAGAACT	2880
TCCACATCCA TAGCTGGAAC TATACGGTCC ATCATTTCGA TAACAGTCAC TTTTGAACCA	2940
AATGTCATGA AGGCCTGACC GAGTTCGATA CCGACAATCT CACCACCGAT GATAACAAGG	3000
CTTTCTGGCA CTTCTGTCAT TTCAAGAATG TCATCACTAG TCATGACAAG TGGAGATTCC	3060
ATACCAGGGA CGTTGATCTT GTTGACTTTT GAACCACCAG CAAGAATGAT TTTCTTGGTT	3120

TCAAGCAATT CAGAACCATT TACCAAGACG TTCTTGTCTT TAGTGATTGT ACCAATTCCT	3180
TTATGAACAG TAACTCCGTA GCTACGAAGA AGTCCTGCAA CACCACCAAC AAGAGTATTA	3240
ACAACTTTAG ATTTAGTTTC TAAAAGTTTT TCCATATCAA CAGTGAAGTT AGGATTTTCA	3300
ATCACGATAC CACGATTTGC AGCATGACCG ATATTTTCAA TAATTTTCAGC GTTATGAAGG	3360
TAGGTCTTGG TTGGAATACA TCCACGGTTT AAGCAGGTTT CACCAAGTTC AGATTTCTCA	3420
ACAAGGGCAA CCTTACCGCC GAATTGGGCA GCTTTAATGG CTGCAACATA ACCAGCAGGA	3480
CCTCCACCAA TCACAACGAT ATCAAAAGCA TCATCGCTCT TACCATCATC GTTTGAGGTA	3540
CTTGCTACAG GTACAGGGCT AGCTTCTGGC GATGCTGCTC CAGCTGTTGG GATGTTTTCC	3600
CTTTCTTCAC CAAGGTAACC GATAACTTCC GTTACAGGGA CAGTTTCACC ATCTCCTTTG	3660
AGAATGGCAA TCAAGTACCC ATCTTCTTCG GCTTCCAATT CCATGCTGAC TTTATCAGTC	3720
ATGATTTCCA AAAGGATTTT TCCTTCTTTT ACAAATTCTC CGACTTTTTT ATTCCATTGG	3780
ACGATTGTC CTTCTGTCAT ATCCACGCCG GCTTTGGGCA TAATTACTTC TAAGGCCATG	3840
TCTTCCTTCC TTTATCTATA TCTTAAAAAT GAATACTCTT GCTCTTAAAT TAACATTGAG	3900
ATTGGCGTTT CAATCAACTC TTTCAAGTCC TTCATAAAT TAGCACCAGC CATACCATCT	3960
ACGACACGGT GGTCAATGGT TAATCCTAAA CTCATGATTG GGCGAATCAC AATTTACCA	4020
TTGACGACAA CTGGCTTCTC GATTGTCGAA CTGACACCAA GGATAGCTGA GTTGGGTTGG	4080
TTAATAATCG GACCAAAGGA CTGAACACCA AACATTCCCA AATTACTGAT TGTGAATGTT	4140
GAATTTTGTA ACTCACTTGG AGCCAATTTA CCATCCAAGG TACGGCCAAT AACATCCTTA	4200
AAGGCTACAA CCAGTTCTGA AAGACTCATC TTCTCAGCAT TGTAACAAC AGGTGTCATC	4260
AATCCATTAT CCATCCCAAC TGCCATGGCA AGATTGACAT AGTTGTGAGT GATAATAGTC	4320
TTGCCATCTT CTGTCAATGA AGCGTTGATG TATGGGTGTT TCATAAGAGT CTTAACAAT	4380
GCAAGCGAAA GAAGGTCTGT TACAGTAGTC TTCTTCCCAG TTGCTTCCAT GATTGGCTCA	4440
AGAACCTTCT TACGAAGAGC CAACATTTC A GTCATATCAA CTTCATAGTT GAGGGTGAAG	4500
GTGGGCGCAG TCAAGTAAGA TTCAACCATG CGTTGGGCAA TAACCTTACG CATTGGTGTC	4560
ATTGGAATAC GTCGATTTT ACCATATGGT GTTACGTTAT CAGGGACTTC TTCCACTTTT	4620
TCAATCTGAG CAGGAGATTT GATGCTATCG TTTTCGATAT TTTTCAGGAAG CAGGGCCAAA	4680
ACATCCTTCT TCATGATTTT ACCACGATGA CCGGTTCTT GGATTTCTTG CCAAGCAATG	4740
TTATGTTTGA GGGCAATTCG TTTTGCAAGT GGCAGAAATGC GAACCACGTT TGTGTCTTTA	4800
TAAGTTTCCA CGTCTTCTT GTGGACACGA CCGTTTGAC CTGAGCCAGA AACGTCGTAG	4860

988

AGGTTTATCC CTAAATCATC CGCTAACTTT CTAGCTGCAG GAGTCGCTCT TAGCTTGTC	4920
TCAGCCATGA CCTCTCCAAT TCTATTTATG ATACAAAGGG CGTCAAAAGC GACTGAAAA	4980
TAGGAAATCG ACGATGGCTT CGATGAAGCC AAGGAGATT ATCTTTTTTC CGATCTTTTA	5040
GCCCGTGCTC TAATCTAAGA TATTAATGAC GAAGAGCTCT GCACCTAAAA GATACAAAGT	5100
TTCTCGTCAG CTTTATTTTA TTTACATAAC TTATCTTATG TAACCCTATT CTTTGTATA	5160
AGTTTTTCGG ATTGCATCTT TGATACTTC AACTGTTGGA ATCATTGCAT TTTCTAGGTT	5220
TTGTGCATAA GGCATCGGCA CATCTTCTCC TGACAACGG CGAATTGGTG CATCTAGATA	5280
GTCAAATGCT TCTGATTCTG AAATAATAGC TGAAATTTC CCGATATAGC CACTTGTTTT	5340
GTGGGCATCG TTGACCAGAA CAACCTTACC AGTCTTCTTC ACTGAGTTTA TGATGATATC	5400
CTTATCAAGC GGAACAAGGG TACGTGGGTC AACAAATTTC ACTGAAATTC CTTCTTCTGC	5460
TAATCTTCA GCAGCTTGAA CCACACGGCG AAGCATTTTT CCATAAGTAA CAACTGTTAC	5520
ATCCGTTCCCT TGGCGTTTGA TTTCACCAAC CCCAAGTGA ATTGTGTAGT CTGGATCAAC	5580
TGGCACTCC CTTTTTGGT TAAATCTGA CTGTACTCA AGTATAATA CTGGGTTGTT	5640
ATCACGGATA GAAGACTTAA GCAGGCCTTT CATGTCCGCA GGTGTTCCAG GTGCCACAAC	5700
CTTAAGTCCT GGAATGTGAG TAAACCAAGA CTCTAGAGAT TGTGAGTGCT GGGCGGCAGA	5760
GCCAACTCCG TTACCAGCTG CACAACGAAC AGTCATTGGA ACCTGACCTT TACCACCAA	5820
CATGTAACGT GTTTTAGCAG CTTGGTTGAC GATATTGTCC ATGGCAATA CAGAGAAGTC	5880
CATGAAGGTC ATATCGACGA TTGGACGAAG TCCTGTCTATG GCTGCTCCTG CTGCTGCTCC	5940
AGAGATGGCA GCTTCAGAAA TCGGACAGTC ACGGACACGT TCTGGACCAA ATTCTCAAG	6000
CATTCCAACA GAAGTACCGA AGTCTCCTCC GAAGACACCG ACGTCTTCTC CCATCAAGAA	6060
CACATTTTCA TCGCGACGCA TTTCCTCAGA CATAGCAAGG ATAATGGTGT CACGGAAGGA	6120
CATTGTTTTT GTTTCCATTT TATCTCTTTC TCCTTAGTCT GCGTAAATAT CTTCAAAGGC	6180
TGATTCAAGC GGTGGGAATG GGCTTCTCTC TGCAAATTTA ACAGAAGCTT CTA CTGCTTC	6240
CTTTACTTGC GCTTGGATTT CTTCCAATTC TTCGGCACTT GCAATGTTAT TTTCAATAAG	6300
GTAATTGCGG AGGTTTTTCA TTGGATCTTT TTGTTTCCAC AATTCCACTT CTTACGCGT	6360
ACGATATTTA CCAGGGTCAG ATGATGAGTG ACCGAGCCAG CGATAAGTTA CACTTTCAAT	6420
CAAGACTGGA CCATTGCCAC TGCGAACATG GTCCACAGCT TTCTGAAATC CTTCATAGAC	6480
ATCGATGACA TTGTTACCGT CTTGATGAA CATTCAGGA ATTCCATAAG CGGCGCTACG	6540
TTGATGGATA TGTCTATAT TGGTCATTTT CTTGATATCC GCAGAGATAC CGTAACCGTT	6600
GTTAATGCAA TAGAAAATGA CTGGCAGGTT CCAGATAGAA GCCATGTTCA CTGCTTCGTG	6660

991

TTTATTCATA GAAATCAACT TGGGTATCCA ACAATTTATC CCCATCATAA ACAAACCTGG	10260
CTGAAAAGAA GGGTTTATCC TCTAAAAGCC ACTCAACAAA GGTGTGGTCA CCTTCCCAAG	10320
TCGGCTTGCT CAAAACCTCA TCATAGGGAA CCCATTCTAG CGTCCCCTCA TTGCAGTCAA	10380
TCAAGTCGCC CTCAAACCTCC GTCACCTTAA AACATAGGT GTACCAGTCT AAATCTGGTG	10440
TAAATTCAGG AAAAGTGATG ACACCTTTTA GAACTGGCTT GGCTTTGAGC CCTGTTTCTT	10500
CAAGGATTTC ACGCGCCGCG CATTCTGGG GCGTCTCTCC TCTCTTAGC TTACCACCCA	10560
CACCAATCCA TTTCCCTTCA TGGACATCAT TGGGTTTCTT ATTACGATGG AGCATGAGCA	10620
GTTCTTTCCC ATTATCAATG TAGCAAATCG TCGCTAACTG AGGCATATTT TCTCCTTATC	10680
TAAGCCAATC GATTGGCTCT TGTCTGTCT CTTTAAAGAA TGCATTGGCC TTGGAAAAGG	10740
GCTTGAACCC CAAAATCCT CTATAAACCG ACAAAGGACT TGGATGGGCT GATTGATAA	10800
TCAAGTGATG AGGATTGGTA ACTAATGCCT TCTTCTTACG TGCATAAGCT CCCCAGAGTA	10860
CAAAAACGAC TGGTCTATCT AGATGATTGA CCACCTGAAT CACAGCATCA GTAAAAGGCT	10920
CCCAGATTTG ACCAGCATGA CCATTGGCCT GTCCAGCAGG AACAGTCAA CAAGCATTAA	10980
GAAGCAAGAC TCCTTGCTCA GCCCAAGCTG TCAAATCATG AGATTCTTCTTA ACTCCGATAT	11040
CATCTGACAA TTCTTTCAAG ATATTTTGCA AGGATGGTGG AGCTGGGATA GAGTCAGGTA	11100
CAGAAAAACT CAAGCCCTGC GCTTGACCTG GTCCGTGATA GGGGTCTTGC CCTAGAATTA	11160
CCACCTTAAC TTCTTCAAGC AGTGTGTCA AGAGAGCCTG AAAAACCTTT TCCTTGGGTG	11220
GATAAATAAT CCCCTGAGAA TAGACCTGCT CCATAAACTG ATTGATTTTC CCGAAATAAC	11280
CCTCAGGTAA TTGCGCCTTA ATCAAAGCAT GCCAAGACGA GTGTTCCATA GCCGACTCGG	11340

## (2) INFORMATION FOR SEQ ID NO: 148:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12127 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

AAAAAATAGA CTTGTTAGAC TATAAATGTA GTAAGCCTAC ACAAGAAAAA TACATAGAGA	60
TAAAGGTGAT TATTATGAAA TTCAAAAAAA TGCTTACTCT TGCAGCCATT GGCTTATCAG	120
GATTTGGGCT TGTGCTGT GGCAATCAGT CAGCTGCTTC CAAACAGTCA GCTTCAGGAA	180
CGATTGAGGT GATTTCACGA GAAATGGCT CTGGGACACG GGTGCTTC ACAGAAATCA	240

992

CAGGGATTCT CAAAAAAGAC GGTGATAAAA AAATTGACAA CACTGCCAAA ACAGCTGTGA	300
TTCAAAATAG TACAGAAGGT GTTCTCTCAG CAGTTCAAGG GAATGCTAAT GCTATCGGCT	360
ACATCTCCTT GGGATCTTTA ACGAAATCTG TCAAGGCTTT AGAGATTGAT GGTGTCAAGG	420
CTAGTCGAGA CACAGTTTTA GATGGTGAAT ACCCTCTTCA ACGTCCCTTC AACATTGTTT	480
GGTCTTCTAA TCTTTCCAAG CTAGGTCAAG ATTTTATCAG CTTTATCCAC TCCAAACAAG	540
GTCAACAAGT GGTCAAGAT AATAAATTTA TTGAAGCTAA AACCGAAACC ACGGAATATA	600
CAAGCCAACA CTTATCAGGC AAGTTGTCTG TTGTAGGTTT CACTTCAGTA TCTTCTTTAA	660
TGGAATAATT AGCAGAAGCT TATAAAAAAG AAAATCCAGA AGTTACGATT GATATTACCT	720
CTAATGGGTC TTCAGCAGGT ATTACCGCTG TTAAGGAGAA AACCGCTGAT ATTGGTATGG	780
TTTCTAGGGA ATTAACCTCT GAAGAAGGTA AGAGTCTCAC CCATGATGCT ATTGCTTTAG	840
ACGGTATTGC TGTGTGGTC AATAATGACA ATAAGGCAAG CCAAGTCAGT ATGGCTGAAC	900
TTGCAGACGT TTTTAGTGGC AAATTAACCA CCTGGGACAA GATTAAATAA AATGTTTGCT	960
CCATAAATCT CTAAGAGAT GCAGACGTTT CATCGTACAA TAAGATAAAG AAGGCAAGTA	1020
GGGAGGTGTC GTATCTCCCT TACTTTCTTC ACTAGAAAGG ACAAGATGTG ACAAACAAG	1080
CCTTCAAAGA AGCAGTTTTT AGGGCAATTT TTTTCATGAG TGCAACAGTA GCTGTTGTAG	1140
CTATTTGCT AATCTGTTTC TTTATTTTTA GTAATGGCTT ACCTTTCATA GCTAACTACG	1200
GCTTTGCCCG TTTTATTATTA GGCAGTGATT GGTGCCAAC GAACATTCCG GCAAGCTATG	1260
GTATTTTACC AATGATCGTT GGTTCCTTAT TAATTACCTT AGGAGCGATT GTGATTGGGG	1320
TGCCAACAGG CATCTTGACA TCGGTGTTTA TGGTTTATTA TTGTCCAAAG CCCGTCTATG	1380
GCTTCTTAAA ATCAGCTATC AACTTGATGG CAGCCATTCC ATCTATTGTT TATGGTTTTT	1440
TCGGCCTACA ATTATTGGTG CTTGGATTA GAAGCTTTTT AGGAAATGGC ATGAGTGTC	1500
TAACCGCTTC GTTACTATTA GGAATAATGA TTTTGCCAAC CATTATCAGT TTGTCAGAAT	1560
CTGCTATCCG AACAGTCCCC AAAACGTATT ATTCTGGTAG CTTGGCTCTA GGAGCTAGTC	1620
ATGAACGGAG TATTTTTAGT GTCATCTTGC CAGTGCGAG ATCTGGTATT TTATCAGCAG	1680
TTATTTTAGG AATCGGTCGC GCAGTAGGTG AAACCATGGC AGTTATTTTG GTGGCAGGCA	1740
ACCAGCCGAT TATTCGAAGT GGAATCTTTT CAGGAACCAG AACCTTAACA ACCAATATTG	1800
TTCTGGAAAT GGCTTACGCA TCAGGTCAGC ATAGGGAAGC CTTATTGCA ACCTCAGCAG	1860
TTCTCTTTTT CTTATTCTC TTGATTAATG CCTACTTTGC CTACTTGAAA GGAAATCAT	1920
CTTATGAGTA AATACCTGCT AAAACTTCTC GTTTATTGTT TTTCAGCTTT AACCTTTGGC	1980
TCTCTTTTT TAATCATGG TTTTATCCTC ATCAAAGGCT TACCTCATCT AAGTCTATCC	2040

989

GAAAAACACCT TCATTGGTCG CACCATCTCC AAAGAAGCAG ACAACGATT TACCGGTATT	6720
TTGCATTTGC TGA CTGAGGG CTGCACCGAC AGCGATCCCC ATACCACCAC CTACGATACC	6780
ATTGGCACCA AGGTTCCAG CATCAAGGTC AGCGATATGC ATAGATCCAC CTTTCCCTTT	6840
ACAGGTTCCA GTGTATTTAC CAAGGATTC AGCCATCAT CCGTTGAGGT CAATCCCTTT	6900
AGCAATAGCT TGCCCGTGT CACGGTGGT TGAGGTAATC AGATCATCTG GATTGAGAGC	6960
TAACATAGCC CCCACGTTAG CTGCCTCTTC-ACCAACAGAA AAGTGCCTCA TTCCTGGCAC	7020
TTTCCCTTC TTTACTAATT GTGCAATTT TAAGTCCATG CGACGGATT CTTCCATCTT	7080
ACGGAACATT TCTAGCAAAA GATTTTATC TAAAGTTGAC ATCTTCTGC CTTTCTAACT	7140
TTCTTCTTAC CTTACTATTT TACCGCTTTT GGCAAATACT GTCAAAGTTT TTCTAAAAGA	7200
AATTTACAA AATAAAAAAG AAAACCCCGT GAAACAAGG GATTTTCTTG TCAAGAATAT	7260
TTTTTCACAA ACTTTTTAGC ATTTGGATTT TGCTAAAGAT TCAAATCTCT TCATAATCAC	7320
AGTTAAACGC CAACGGTAGA GCGCCCCGCT CACAATCAA CTAATAATCA AGCCGATCCA	7380
GTAAGAATAA GCTCCAAAAT CTGTTAGGGA ATCAAATAGC GTAAACACAGG GATTGCTACG	7440
CCCCAATAAC CAAGCAAACC AAGGTAAAA GGAATAACTG TATCCTTATA CCCCCGAAA	7500
ATTCCTGAA GCGGCGCCGC AAAGGTATCT GCTAACTGGA AGAAAAGACT ATAAGTAAA	7560
AAACGCACTG TCAAATCGAT AAATTTTGGG TCGTTACCAT AAAGACTGGC CACATTTCCC	7620
CTAAAAATGT AAAGGAAGGT TAAGGTGAAG GCCGCAAAA TGAGGGCAGT CCATCTTCCT	7680
AGACCAATAT AGGTTTTCGC ATCATCAAAT CGCTTGGCTC CCACTTCATA GGAAACGACA	7740
ATAGCCATAG CCGATGAGAT ACTCATAGGA AAGGCGTACA TAAGACTTGA AAAGTTCATA	7800
GCTGACTGGT GACTAGCTAT AATCAAGGGC GAAAACCTAG CCATAATCAA GCCAACCCT	7860
GAAAAGATAG CCACTTCCGC GAAGACAGTT CCCCCAATAG GCAGACCTAA ACGAACTCCT	7920
TCCTTAATTT TATCCATATT AAGTGAATT CGTTTCTCAA GGTGTAAGGC TTTGAGCTTC	7980
TCCTGTTTAA ATAAAACCAG AACAGAAATC CCAAGCAAGA CCCAGTAGGC CAAGGATGTT	8040
CCTAAACCAG CACCAGCCCC TCCAGTTCT GGAACACCAA AGGCACCGTA AATCAAGAGA	8100
TAGTTAAATC CGCTATTGAG AGGGAGTAAC AAAAGCATGA GGTACATGGA CAGTTTGGTC	8160
AAGCCCAGCG AATCCAGCAA GGAACGAATG ACGCTAAAGA GCAACAAGGG GATAATCCCC	8220
ATAGATAAAA ACCAAAGATA GCGAACCGCT ACTGCCGCTA CTGCTGCTTC TAACCCAATA	8280
TGATTCAAGA TTATTGGTGC CAAGAAAAGT ACCATCCCCA GCAAGACCAC AGATAGGCCC	8340
AAGGCCAAAT AAATAAATTG GTAAAAATCA GACGCAACTT CTTCTTTTTT GCCTCGACCA	8400

990

AGATGGTGAC CAATGATAGG CACCAAGGCT GACACAATCC CTGTTAGAAA TGTAAGAGAA	8460
GGATTCCAGA TACTGGTTGC CATAGATACA CCAGCCAAGT CCATAGTGTT GTATTGACCT	8520
GTCATTGCAG TATCAACAAA AGAGGCAGAA TAATTGGCAA ATTGGTAGAT CAGGATTGGG	8580
AAGAAAATTT TAAAAATAA TACTAACTTC TCTCGTAAAC ACTTTGTCTT ATACATACTT	8640
CTCTTTCTAT TCTGATTTAT CTAAACCAAA GAGTTTCAGA CCATAGTTTT TCAAACCTAG	8700
CGGAGGTTTA TTAGATTTTG AAGTAGTATG CCAACACGCA CATGTACGAC AATAATAGCT	8760
TCTAACTAAA CCTCCGTTAT CATATTGAAC CGCATGGTCA GCTTTTCTT TAGTTTCATA	8820
TTGAATTTTG GAACGATTAG CTGCGGGACA GTAAATTCCA CTATTAGATT TCGCTTGTCT	8880
CTCCCTACGT TTTGAAAAAT AATTCATATT CTAACCTCTA TCAAGCTTGA TAGACGATTT	8940
GTCCCTTACA GATGGTATAT TTAACCTGCC CTTTAAAGGT TTCACCGATG AATGGTGAAT	9000
TAGCTGCTTT GGAAGCAAAA TGGGAGTCCA CAAAGCGGTC AGCCTTGGCA TCAAAAATAG	9060
TGATATCTGC TGGACCATTC TCAGCCAAGT AACCTGCTTC AAAGTTGTAA AGCTTGGCTG	9120
GGTTGTATGT CATTTTTTCA AGTAATTCCA TCAAGCTCAA CTCACCAGCT TCTACTAAAT	9180
AGGTCAAGCT GAGAGACAGG GATGTTTCTA AGCCAGTCAT ACCAGATGGC GCTTTGGTAA	9240
TATCCTCAAC ATTTTTTTCA TCTACATGAT GAGGCGCGTG GTCAGTCGCA ATAAGTGTGA	9300
TGACACCTGA TTTGAGACCT TCGATAACGG CACGACGGTC TGATTCCAAA CGAAGCGGTG	9360
GATTCATCTT AGCATTGCTA CTTGTGTGTA AAAGAAGTGC TTCTGTCTTA GAGAAATGCT	9420
GTGGCGCTAC TTCTGCTGTG ACTTCTGCAC CTAACCCCTG AGCAAACCTC ACTACTTTAA	9480
CACTTTCTTC CTTAGACAAA TGCTGGATGT GAACATGGGC TTTAGTTGCA TAGGCAATCA	9540
TGACATCAG CGCCATCATA GCGTACTCAG CCACCCAGT AGCACCGCAG ATATGGAAAT	9600
GTTCTCTAGC AATATTTTCA TTAAAGCCAA GAACACCGTT CAAACCTGGA TCTTCCTCAT	9660
GAAGGCTGAT AAAGGTATTG AGTTTTTTGG CTTCTCCAT GGCTTCCTTG ACAATCTTAC	9720
TGCTCTCAAG CGGAATACCG TCATCAGAGA AACCAACCGC ACCAGCTTCT AAGAGTGCT	9780
TAAAGTCAGT CAAGTTTTTA CCATTAAAGT TTTTAGTAAT GGTGCAACT GTCTTGACAT	9840
TAATCTTCTC TTTGGCAGCT GACTGGAGAA CTGCTTGCAA AGTCTCCAG TCTGAAATGG	9900
TTGGACTGGT ATTAGCCATC ATGACGACAG TAGTAAACC ACCTGCAGCG GCTGCTAGGG	9960
CACCAATATG AATGTCTTCT TTATGTGTTT GACCAGGTTT ACGGAAATGA ACATGAATAT	10020
CGACCAAGCC AGGAGCAACC ACAAGACCAG TAGCATCAAT CGTTTCTGCT CTTCTTCCG	10080
TGATCTCAGA CGCAATTTTG ATAATTTTCC CATCTTGAAC TAAGACATCA CAACTTGAT	10140
CCAAACCAGA CTTGGGATCC ATTACACGAC CATTTTTGAT TAGTAGCATC TGCTTTCTCC	10200



993

CTCTTTTCTT GGACTTATAC TTCTGAGAAC ATTTCCCTTA TGCCAGCGAT TATTTCCACC	2100
GTTATTCTGG TCTTTGGTGC TCTTCTTTTA GCCTTGCCCA TAGGGATTTT TGCTGGTTTT	2160
TATCTTGTGG AATATACAAA AAAAGATTCC CTTTGTGTTA AAATCATGCG ATTGGCCTCA	2220
GATACCTTAT CTGGGATTCC TTCCATTGTT TTTGGTCTGT TTGGCATGCT CTTCTTTGTA	2280
GTCTTCTTAG GTTTTCAATA CTCTCTGTTA TCAGGAATCT TAACCTCAGT TATCATGGTG	2340
TTGCCAGTCA TTATTCGCTC AACAGAAGAA GCCCTTTTAT CTGTTAGTGA TAGCATGCGT	2400
CAAGCAAGTT ATGGACTTGG GGCAGGTAAG TTACGGACTG TTTTTAGAAT TGTTCTACCA	2460
GTGGCCATGC CAGGTATTTT AGCTGGAGTG ATACTAGCTA TTGGCCGTAT CGTTGGTGAA	2520
ACAGCTGCCC TCATGTATAC ATTAGGTACC TCTACCAATA CGCCAAGTAG TCTCATGTCT	2580
TCAGGCCGTT CTCTAGCCCT ACATATGTAT ATGCTGTCAA GTGAGGGGCT ACATGTCAAT	2640
GAAGCCTATG CTACCGCGT GATTTTGATT ATTACTGTTT TAATGATAAA TACTCTATCA	2700
AGCTTATTAT CTCGAAAAC TGTGAAAGGA GCTTCCTAGT ATGGGAACAT TTTCAGTCAG	2760
ACACCTAGAC TTATTTTACG GGGATTTTCA AGCCTTAAAA AATATTTTCA TTCAATTACC	2820
AGAAAGACAG ATTACTGCCT TGATAGGCCC ATCTGGTTGT GGCAAATCAA CTTTTCTAAA	2880
AACCCTTAAC CGGATGAACG ATTTGGTTCC TTCTTGCCAT ATTGAAGGCC AAGTCCTCTT	2940
AGATGAGCAA GATATTTATA GTAGCAAATT CAACCTTAAT CAGCTACGTA AGCGTGTAGG	3000
GATGGTTTTT CAACAGCCTA ATCCCTTTCG CATGTCTATC TATGATAACG TGGCTTATGG	3060
CCCAAGGACA CATGGTATTC GAGACAAAA ACAATTAGAT GCCTTAGTGG AGAAATCTTT	3120
AAAAGGGGCA GCCATTTGGG AAGAAGTCAA AGATGATCTT AAAAGAGTG CCATGTCCTT	3180
ATCTGGCGGT CAGCAGCAAC GCCTTTGCAT TGCGCGAGCT TTAGCAGTAG AACCTGATAT	3240
TCTGTTAATG GATGAGCCGA CTTCAGCCTT AGACCCTATC TCCACTTTAA AAATTGAAGA	3300
CCTCATTGAG CAACTAAAA AGGATTATAC GATTATCATT GTTACCCATA ACATGCAACA	3360
AGCTTCACGT ATTTGAGATA AAAGTCTTT TTTCTTAACA GGAGAAATTT GCGAATTTGG	3420
AGATACCGTT GACGTGTTTA CCAATCCAAA AGATCAGCGC ACAGAAGACT ATATTTGAGG	3480
ACGGTTCGGA TAAGGAAGGA AAAACCTATG AGAAATCAAT TTGACTTAGA ATTGCATGAA	3540
TTAGAACAAT CCTTTTTAGG ACTAGGGCAA CTGTCTCTG AAACAGCTTC AAAAGCCTTA	3600
CTGGCCTTAG CCTCCAAAGA CAAGGAGATG GCAGAGCTAA TTATCAATAA GGATCATGCT	3660
ATCAACCAAG GTCAAAGCGC TATCGAATTG ACCTGTGCCC GTTTGTTGGC CTTGCAGCAG	3720
CCACAAGTGT CTGACCTTCG ATTTGTGATT AGCATCATGT CTTCTTGTTT AGACCTTGAA	3780

994

CGTATGGGAG ACCATATGGC AGGCATTGCC AAAGCTGTTT TGCAACTAAA AGAAAATCAA	3840
CTAGCCCCCTG ACGAAGAACA GTTACACCAA ATGGGTAAAT TATCCCTCAG CATGCTAGCC	3900
GATTTATTGG TTGCCTTTCC TTTGCACCAA GCCTCAAAG CTATTAGTAT TGCTCAAAAA	3960
GATGAACAGA TTGACCAATA TTATTATGCC TTATCAAAGG AAATCATTGG ACTTATGAAA	4020
GACCAAGAAA CCTCAATTCC CAATGGAACCT CAATACCTTT ATATCATAGG GCATCTGGAA	4080
CGCTCGCTGA TTACATTGCT AACATTGTG AACGCCTAGT CTACCTAGAA ACAGGAGAAC	4140
TAGTGGATTT GAATTAATTC AACTAATCCT TAAAAGAGAA GAGTACGATT AAGTACTCTT	4200
TTTTATGGTT GTAAAAAGT TCATTTGACC AATTTAAGCA GTGTAGATAG TGAGGAGTTG	4260
TTTCAATTCT ATCGTGAACG AGGGAATGCT GAAACTTTA TCAAAGAAAG GAAAGCAGGA	4320
TTCTTTGGGG ATAAGACAGA TAGTTCGACC ATGATTAAGA ATGAAGTACG TATGATGATG	4380
GGCTGTCTGG CTTATAATCT CTACCTCTTT TTAAGCAGC TAGCTGGTGA TGAAGTAAAG	4440
TCCTTGACTA TCAAGCGTTT TCGACGTCTC TTCCTTCATA TTGCCGAAA ATATGTCTCT	4500
ACTGCTAGAC GACATATTCT CAAATTCTCA AGTCTATACG CCTATTCAA ACAGTTTCAA	4560
GCCTTATTTG ATACAATCTG CCAGATAAAT CTGATACTCC CTGTTCCATA TAGAGCTAGA	4620
GGGCAGGGGA AAACATGCCT AACAGAATAA GTCACCTTAT TTTAAAAATC GAGCATCAAA	4680
CCAAGGGAGG AGTCTGCCCT TTTTAGGAA AAAATCAAGA CAAATCTCCT CAATTATGTC	4740
TCGAACATCA GAAATTAAGC AAAATCACCA GAAGGACAGT ATTTCAACTA GCTTTTCTGG	4800
TAATTTTGA ACTGTGTAGT TCGTTAGTGC CAGATATGAA TAATTTGGGA TGATAAATCT	4860
TTCTTCTCA GGTAGCCTAT CATAATACTC TTCAAAAATC TTATCAAAAA CACTCTCTTT	4920
CTTTTGGGCG ATAGTTTCAT CTTCGTATGT AGGAGTCCTC ATCAAGAAAT ACTTCAATTC	4980
TAGGTATTCC TTATCCAACCT CTATATAACT TGGCATCAAC TTGTAATCTT CAACCCCAA	5040
ACGTTCAGCA ATATATTTTA ACTTTGTTAG TATTGGTCTG GATTCTCCAT TTTCAATTCT	5100
AATTAATTGA CGGATACTTA ATTCAGACTC ATCACCACAA AATTCTGAAC GACTGATTCT	5160
TTTAGCCAAA CGTAATCTTT TAATTTTTC GCCAAACTCT CGCAACCTAC AAGAACTTCC	5220
TGAGTTGTTT ACCTCTATTA TAAGCATATA CTGAATCAAA CTATCTATCA GATTTCTTCT	5280
CACTTTAACT AAAGACTAAG AGTTTATCCC TTCGTCTCGG TTTTGTGTA TTTTCCACC	5340
ATACCCAGT AATGCAAGTG CAAAATCCCC TAGAATATGA TAGAATAAGA GAAAGAACTC	5400
TATCAAGGAG GAAATCATGG AAAACAAAC CGTCGCCGTC TTGGGGCCTG GTTCTTGGGG	5460
AACCGCCCTT TCACAAGTCT TAAATGACAA TGGACACGAG GTACGTATTT GGGGAAATCT	5520
TCCCGAGCAA ATCAATGAAA TTAATACACA CCATACTAAT AAGCACTACT TTAAAGATGT	5580

995

CGTTCTAGAC GAAAATATCA TTGCCTACAC CGACTTAGCA GAAACATTGA AAGATGTGGA	5640
TGCGATTTTG TTTGTTGTCC CAACAAAAGT GACACGACTT GTTGCCCAGC AAGTTGCACA	5700
AACCTTGGAC CATAAGGTTA TCATCATGCA CGCATCAAAG GGATTAGAAC CTGATAGCCA	5760
TAAACGATTA TCAACCATTG TTGAAGAAGA AATTCCTGAA CATCTCCGTA GTGATATCGT	5820
CGTTGTTTCA GGGCCTAGTC ATGCAGAAGA GACCATTGTG CGTGACCTAA CTTTAATAAC	5880
TGCTGCTTCT AAAGATTIAC AAACAGCTCA ATACGTTTCA AAGCTATTTA GTAATCACTA	5940
CTTCCGACTT TATACCAATA CGGATGTTAT CGGGGTTGAA ACTGCTGGTG CTCTTAAAAA	6000
TATTATTGCT GTCGGTGCTG GAGCTTTACA TGGTCTTGGA TTTGGTGATA ATGCTAAGGC	6060
AGCCATCATC GCTCGAGGTT TAGCAGAAAT CACCCGCCTA GGGGTAGCAC TCGGGGCCAG	6120
TCCATTGACC TATAGCGGCT TATCTGGTGT GGGAGATTG ATCGTAACGG GAACTTCCAT	6180
CCACTCTCGT AACTGGAGAG CTGGAGATGC TCTCGGACGA GGAGAATCCC TAGCTGATAT	6240
AGAAGCTAAT ATGGGCATGG TAATCGAAGG AATTTCAACG ACTCGAGCAG CCTATGAACT	6300
AGCCCAAGAA CTTGGAGTCT ATATGCCCAT TACACAGGCT ATTTACCAAG TTATTTATCA	6360
CGGAACCAAT ATCAAAGATG CCATTTATGA CATCATGAAC AATGAATTTA AAGCAGAAAA	6420
TGAGTGGTCT TAACCCTCTA TAGAAAGGAT TTTTATGACA TCAAAAGTTA GAAAGGCAGT	6480
CATCCCTGCT GCTGGACTAG GAACTCGATT TTTACCAGCA ACCAAGGCCC TTGCCAAAGA	6540
AATGTTGCCA ATCGTAGACA AACCAACTAT CCAGTTTATC GTGGAAGAAG CTCTCAAATC	6600
AGGTATTGAA GATATTCTAG TTGTCACTGG TAAATCAAAA CGTTCTATTG AGGACCACTT	6660
TGATTCAAAC TTCGAATTGG AATATAACCT CAAAGAAAAA GGGAAAACAG ATCTTTTGAA	6720
GCTAGTTGAT AAAACAAC TG ACATGCGTCT GCATTTTATC CGCCAAACTC ATCCACGCCG	6780
TCTCGGAGAT GCTGTTTTGC AAGCCAAGGC TTTCGTCGGA AATGAACCTT TTGTCGTTAT	6840
GCTTGGTGAT GACTTGATGG ATATCACAGA CGAAAAGGCT GTTCCACTTA CCAAACAAC	6900
CATGGATGAC TACGAGCGTA CCCACGCGTC TACTATCGCT GTCATGCCAG TCCCTCATGA	6960
CGAAGTATCT GCTTACGGG TTATTGCTCC GCAAGGCGAA GGAAAAGATG GTCTTTACAG	7020
TGTTGAAACC TTTGTTGAAA AACCAGCTCC AGAGGACGCT CCTAGCGACC TTGCTATTAT	7080
CGGACGCTAC CTCCTCAGC CTGAAATTTT TGAGATTCTC GAAAAGCAAG CTCCAGGTGC	7140
AGGAAATGAA ATTCAGCTGA CAGATGCAAT CGACACCCTC AATAAACAC AACGTGTATT	7200
TGCTCGTGAG TTCAAAGGGG CTCGTTACGA TGTCGGAGAC AAGTTTGGCT TCATGAAAA	7260
ATCCATCGAC TACGCCCTCA AACACCCACA AGTCAAAGAT GATTTGAAGA ATTACCTCAT	7320

996

CCAACTTGGA AAAGAATTGA CTGAGAAGGA ATAACAAAAT CATTTATATA AAGATTAGCC	7380
ACACATAAAT TAAGTAAATT CTCTACTTGA ATCTACCTAT TTAATAAAAA CTAATGAAAA	7440
CGCTATACTT GTATTTGTTT TTTCATTAAA ATAAGAGTAG AATAAATTAG TATAGTAAAA	7500
CAAAAAAGCA CCGAATCGGT GCGCACTTTT TCAAGTTGTG TACCGACAAA GCCTTATTTT	7560
AACTTTGCTA TGTGTGTTCT AATGGTTCCA AAATAATAAA TATTTTAA TTTGACTTAA	7620
CTGTTGGAGT AGTCATGGTT AAATTAAATC AACCGAGCCG AACATAAGTT GTTTAATTTT	7680
GTGGAAGCTA TTAATAAAAA TATAATAAGG GAGAAAGATA GGTGTAATTT TAATTTTAAA	7740
GTAATTGCGG ACACTATCAA AGAAAAAGAT TATGGAGAAC AAATTTGTAG AATTTATCGA	7800
AAACAATAAA AAAGTAATCA TTTCATCAGT TGCAGTTGGT GTTGTATTGG TATTAGGGTT	7860
TGGATGGTAT TCATATAACC AACACAAGC AGAACAACAA GCAAAAATTG TACAATTAGA	7920
AAAAGATAGC AAATCAGACA AAGAACAAGT TGATAAACTA TTTGAATCAT TTGATGCATC	7980
TTCAGATGAA TCTATTTCTA AATTAAAAGA ACTATCTGAA ACTTCACTTA AAACCGATGC	8040
AGGTAAAGAC TATCTTAATA ACAAAGTCAA AGAATCATCT AAAGCAATTG TAGATTTTCA	8100
TTTGCAAAAA GGTTTGGCTT ATGATGTTAA AGATTGAGAT GACAAATTTA AAGATAAAGC	8160
AACCTTGAA ACAAATGTAA AAGAAATTAC AAAACAAATT GATTTTATCA AAAAAGTTGA	8220
TGAAACTTTT AAACAAGAGA ATTTGGAAGA AACTCTTAAA TCTCTAAATG ATCTTGTTGA	8280
TAAATATCAA AAACAAATCG AACTTTTGAA GAAAGAAGAA GAAAAAGCTG CTGAAAAAGC	8340
TGCTGAAAAA GCAAAGGAAT CTTCTAGTCA AAGTAATTCT TCTGGTAGTG CTTCTAATGA	8400
GTCTTATAAT GGATCTTCCA ATTCAAATGT AGATTATAGT TCATCTGAAC AACTAATGG	8460
ATATTCAAAT AATTATGGCG GTCAAGATTA TTCTGGTTCA GGAGATAGTT CAACAAATGG	8520
TGGATCATCA GAACAATATT CATCTAGCAA TTCAAACAGC GGAGCAAATA ATGTCTACAG	8580
ATATAAGGC ACTGGTGCTG ACGGCTATCA AAGATACTAC TACAAAGATC ATAATAATGG	8640
AGATGTGTAT GATGACGATG GAAATTACCT TGGGAACCTT GGTGGCGGCA TTGCAGAACC	8700
TAGTCAACGC TAATAACTAT TTTAGAGCTG TGTTGTTTCG AATGGTTCCA AAACACATTA	8760
AAAGCTACTC ATTTTTTAAG TAGCTTTTTT CTTATTCAAG TTTACATATT ATACTCAATG	8820
AAAATCAAAT TCAAACCAGC TCAGCATCGC CTTACCGTAG GTATGGTTAC TGACTTCGTC	8880
AGTTTCATCT ACAACCTCAA AACCATGTTT TGAGCTGACT TCGTCAGTTC TATCTACAAC	8940
CTCAAAGCAG TGCTTTGAGC AACCTGCGGC TAGCTTCCTA GTTTGCTCTT TGATTTTCAT	9000
TGAGTATTAG TCGTCACAAT CCCATTCCCT TGTAGAAAAG CAAAATGGCG AGTCCTACGA	9060
ACAAGACTAC CGCTCCTAAT CTCTGGCTGG TGTATACAT CCGTTTTTCT CCTCTAACTG	9120

GAAAGATAAC TGCTAGAAAT GCGCCACCAA CTGCACCACC GATATGGCCT GCTAGGCTGA 9180  
TTCCTGGAAT CAGAACACTT CCAATAATGT TAACCACAAA AAGTGTGAGA TAGGATTGCC 9240  
CTAGCTGTTG GATATAAGGA TTGCGAGTTG CATAGCGAAG AACAAATAATC GCGGCAAATA 9300  
GCCCATAAAG AGAGGTAGAG GCGCCTGCTG CTAAGGATTT AGGACTAAAT AAAAAACAA 9360  
AGAGATTGCC CATCATTCCT GATAAAGAT AGAGAAAGAA AAAGTGTCTTA GAACCGAAAA 9420  
TCTCCTCTAC CTGCCTTCCA AGATAATAAA GTGAAAGCAT ATTAACAATG AAATGTTCCC 9480  
ACCCAATATG AACAAAAATG GCAGACAAGA GACGCCAAAC CTGCTCGGGA AAGAGCGGAA 9540  
TAGCTGGCCC ATACATGGCT CCAAAATCGAA ATAATGTATC TGCCCTGTCA AAGTTTCCGC 9600  
CTGCAGTGAC CAACATTAGT AAAAATACCA AGGCCGTCAC TAAGAGGAAG AAAGTGTCTA 9660  
CAGGGTAACG TCTATCAAAG ATTTCTTCA TCAATTAATA CCTCCTGAAC AGGAATATCA 9720  
TGGTTTTCAG GTATAAAGTC CTGAATTTGA CAAGGATATA TCGTACTCAA AGTACGACCA 9780  
GAAAAATGTT CCAGATAGCG GTCATAATAG CCTCCACCGT ATCCTATCCG ATATCCTTTC 9840  
GTCGTAAGG CCAGACCAGG AACATGAATC AAATCAATCT GAGATGCATC CACCACTTCC 9900  
AAATCTCCCT GTAGCTCCAG TAAGGCAAAG AAAGTTTTTA CCAACTGTTG CGGATCATAG 9960  
ACCACAAAGT CCATGCGCCC CTTGGGATAA GTTTTGGGTA TTAACCTT CTTGCCGTCC 10020  
TTCAGCGCCT GCTCAATCAG TTCCTGCGTT TGAACTCAT GAGAAAAAGA GAGGTAGGTT 10080  
GCGATGACCT TGGCTTCTTG ATAAAAGGGG TGTGTAAAA GCGCTCGGT TAAAGCTTGG 10140  
TCTATAGCCT GTTTTGTCTC TTGAGATATA GCCTTCATTT CATGCAAGAC TTGCTGCGT 10200  
AATTCGATT TCATAGACAA GCCCTCTATT CTGCTGCCTT CTTTTTCAGG AAAGTAGACA 10260  
CCGCAGCCAC CCCAATAGCT AAGACTTCTT CCTTAGGACT CATTTGAGGG TGATGAAGAG 10320  
CGTAGGGACT ATCGATACCT AGCCAAAACA TCACGCCATC AACCTTTGAA AGGAGATAAC 10380  
CAAAGTCCTC GCCTGTCATA GCAGGTTGGA TATCAATCAA CTCGATTCCG TCTTTTTCGT 10440  
CAAAGAAGTC CATCAGTTCA CGCGCAAGG CTGGATTGTT CTCAACAGGT AGGTATCCAC 10500  
CTTGTTTGAG TTCCACTTCG ACTTCCATAT CAAAGGCAGC TGCAACCCCT TCTGCAACTG 10560  
TTTTTACCCT CTTTTCGACC AAGAGACTCA TGTCTGTGT CAAGGCACGA ATAGTTCCAT 10620  
GTAAAAAAGC TGTGTCTGTG ATGACATTGT TGGTGGTTCC AGCTTGAAAA ACGCCGAAGG 10680  
TCACCACTGC TCCCTCGATT GGGTTGACAT TGCGGCTAAC AACTGACTGC ACTTGGGTCA 10740  
CAAAGTAACT AGCCGCCACC AAGCGTCAT TGCTTCATG AGGAAAAGCT GCGTGGCCAC 10800  
CTTTCCTTT GAAACGGATC TTCACCTCGC AAGTTCCTGC AAAGAGTGTA TGAGTATTAG 10860

998

TCGCAATCTG GCCGACTTTC AAATCTGGAC GAACATGGAG ACCATAGAAT TGATCTGGCA	10920
ACCAATCTCC AAAAGCACCG TCCTCATACA TGAGCATACC ACCAGCTTCA TTTCTTCAG	10980
CAGGCTGAAA TAGAAAGAGC AGATTATTCT TGGGTGCTC CTCAAGGGCG CGCTCAAGAC	11040
AGCCTAAGGC AATGGTCATA TGAAAATCAT GGACACAGGC ATGCATGCGA CCTTGGTGT	11100
GAGAAGCAAA AGGTAGACCT GTTTGTTTGA CGATAGGCAG GCCATCAATA TCTGTCCGCC	11160
AACCAATGGT TCGCTCCGGC TGAATCCCT GCAGGTAGAC CAAAATCCCT GTCCGCCAAG	11220
TACGAATTTG AACAAAATCC TTGCCCGTAG TCAATTTCTC AATCACATCC AGCAAATAAG	11280
CCTGAGTCTT GAACTCCTCC AAGCCAATCT CTGGAATCTG GTGTAAATCT CGTCTAGTCT	11340
GAATCAAATC TAACATCTAT CTGTCTCCG ATATAGCAGA AAGAGGCTGG AAAAAGGGTT	11400
CCGCCTCTTT TTTACTTTTA CAATTACAAG GTACGAAGCG CATCCTCTAG CGCTGTTTTT	11460
TGTTGAGTTT GGGCATCAAT TTCTTTGATA ATACGAGCTG GAACACCTGC TACTACCACG	11520
TTTCTGGGA CATCTTGGGT AACAATAGCT CCTGCTGCGA CAACTGAACC ACTACCGATT	11580
TGGACTCCTT CGATAACCAC TGCATTAGCA CCGATAAGAA CATTGTCTCC GACACGGACT	11640
GGTTCAGCAC TAGCTGGCTC AATCACACCT GCCAAAATG CACCTGCACC AACGTGGCTA	11700
TTTTTTCCAA CGATGGCAGC GCCACCAAGG ATGGCACCCA TGTCAATCAT GGTTCAGCA	11760
CCGATTTTCA CACCGATATT GATAACAGAT CCCATCATGA TAACAGCATT GTCACCAATT	11820
TCCACCTGGT CACGGATAAT CGCACCTGGC TCGATACGAG CGTTGATAGC ACGCTTATCT	11880
AGCAAAGGAA CTGCAGAATT ACGAGCATCT TGCTCGACAA CATAATCTTG ATTTTCTACC	11940
AAACCTTCAA GAAGCGGAGC CACATCCTC CAGTCTCCGA ATAGGACATT TCCTAGTTTG	12000
ACAACAGAGC TAGGCACAGC AGTTGCGAGT TGCCCTCAA AGGTTACTTT GAACTGGTT	12060
TTCTTTTTCAG CATTGGCGAT AAATTGGATA ATTTCTTGAG CGTTCATTTT TGTAGCAGTC	12120
ATAGGTG	12127

## (2) INFORMATION FOR SEQ ID NO: 149:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 12566 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

CCATCCTTCT GTTGATGTGA CAGGAATGAT GATAAATCAA CCAGTAGCTA GTCGCGAAGA	60
GGTGACAGAG GCTTTGAGTC ACTTGCGCGT AGAGCACAAAT AGTCTCATTG CTCGTCGAAT	120

999

CGTTGAGCCA AATGAAGCTG GAGAAACACG CTTTACCTAT GCCACTTATG GTGAGGGAAA	180
GCTTCCAGAA GGTCTGACCA TTTCTCTCAA GGAGAGTGCA GAAACGAGTG ATTTATTAGG	240
GTCTTACTTG ATTGTATCAG GAAGTTTGGG TGGAGTGAGC TTACAGACCA CCTTGAAAGA	300
GCTTGGTTAT CAAGGCTTTG TTTTGAATGG AGAAGATCCA TTTTCGATAG TCTTACTATT	360
GACGGCCACC CCTATGGTGC TACTGAGTTT AGCTATTTT CTGCTGACCT TTATGAGTCT	420
GACCCTGATT TATCGGATCA AATCCCTTCG TCAGGCAGGG ATTCGCTTAA TAGCTGGTGA	480
GAGCTTGTTT GGAGTTGCTC TCAGACCAGT GTTAGAAGAT GTGAGACAGC TTATCTGCTC	540
AGTGCTGGTA TCCAGTCTTT TGGGATTGGG GATTCTCTGG TATCAAGGTG CCTTGTTTAT	600
GGCAACGGTG CAACTGGTCA TCATTGCTCT TCTACTTTAT GGATTGACCT TGGCAGGGAT	660
TTCTACCTTA CTAAGTGTCG TCTATCTACT TGGTTTACAG GAAAATAGTC TGGTGGATCT	720
ATTGAAAGGG AAACCTCCCTC TCAAACGTAT GATGACATTG ATGATGGTGG GGCAACTCTT	780
AGCTGTATTG GTGGTCGGAT CGAGTGGGAC AGCTCTCCTA CCCCCTACC GTGAAATGCA	840
GGAAATGGAG AGAGCTAGCA ATAAATGGAG CCAGTCTCTA GACCGTTACC GTCTATCCTT	900
TGGTTGGTCT AGTGCATTTG CCGATGAAGA AGGAACGCGT AAGGATAATC GTGAGTGGCA	960
GACATTTACT GAAGAACGGT TAGCCAATAC AGACTCTTTT TATATTATGA GCAATGTTGA	1020
CAATTTCTCA GATGGAGCAG AAGTGGACCT AGATGGCAAT CGTCTCAGTG ACTACACACC	1080
GTCAGGGAAT GTTATCTATG TCTCACCCTG CTATCTGATA GAACAAAAGA TTACCGTTTC	1140
TTTCAGATTT ATGGACAAGA TGCAAACTT GTCTGAGGGA GAGTTTGGGC TGATCTTGCC	1200
TGAGAGCTTG CGAGAGCAGT CTGTCTACTA CCAAGGATTG TTTACAGATT ACCTGCAAAA	1260
CTTTTCATCT GAAAGTGTAG AAGTGACGAG TCAGAAACAC TACCTCCCAC AGGTAAGGCT	1320
AGCTTTTACA GAAACAGGAC AGGAACGTTT CCTCTATAAT GATGGGTACA AGACAACACG	1380
CCAGTACCTA AAAGATCCGA TTATTGTAGT TCTAACGCCG CAAGCGACTG GAACAAGACC	1440
TGTTGCAGGG ATGTTGTGGG GAACTACGGC TAATAGTGCC TTGAACTAG ATCGATATGG	1500
AGACAGCATC ACAGCTCTAA AAGAGAAAGG TCTGTATCAC AAGGTTTCTT ACTTGGTAAA	1560
AAGCCAGCTA TTTTTTGCCA AGGTACTAAA TGACAAACGG GTGGAGTTTT ACTCTCTCCT	1620
TATTGGGACG ATTTTGACCC TGTCTACGGC TATCTTGTTA TTTGATTCCA TGAATCTTCT	1680
CTATTTTGAG CAGTTCAGAC GGGAACTTAT GATTAAACGT CTTGCTGGTA TGACAATCTA	1740
TGAGCTTCAT GGCAAGTATT TACTGGCGCA AGGAGGAGTT CTCTTGCTTG GCCTAGTCCT	1800
ATCTAGTATT TTGACAAGAG ATGGTTTGAT TAGCGCTCTA GTTGTAGCTT TGTTPACGCT	1860

1000  
TAACGCCCTC TTGATTTTAG TAAGGCAGGA CAAAAAAGAA GAAGCTGGTA GCATGGCAGT 1920  
ATTGAAAGGA AAATAAGATG ATTGATATTC AAGGATTGGA AAAGAAATTT AATGACCGCG 1980  
CGATTTTCTC TGGTTTGAAT CTCAAGCTGG AGAAGGGCAA GGTATTATGCC TTAATCGGAA 2040  
AGAGTGGAAG CGGAAAGACG ACGCTGCTGA ATATCTTGGG AAAGCTAGAA AAGATAGATG 2100  
GTGGAAGGGT TCTCTATCAG GGGAAAGATT TAAAAACCAT TCCCACTCGT GAGTATTTTC 2160  
GAGACCAGAT GGGCTATCTC TTTCAAAATT TCGGCCTCTT AGAAAACCAA TCAATCAAAG 2220  
AAAAATTTGA TTTGGGTTTT GTTGGTCAGA AAATCTCAA AGTAGAACGT TTGGAAGGC 2280  
AAGTGGGGGC TTTAGAAAAA GTTAATCTAG GGTATTTGGA TTAGAACAA AAAATCTATA 2340  
CTTTATCTGG GGGAGAGGCC CAACGAGTTG CCCTTGCTAA GACTATTTTG AAAATCCAC 2400  
CCTTGATTTT GGCAGATGAA CCAACAGCAG CTCTTGATCC TGAAAATCA GAGGAGGTTA 2460  
TGAATCTCTT GGTGGATTG AAAGATGAAA ATCGAATTAT CATCATGCG ACCCATAATC 2520  
CCCTAGTCTG GAATAAGGCT GATGAAATCA TTGATATGAG GAAACTTGCT CATGTGTGAA 2580  
AAAATCCGTA TTCGCAGGGT ATCTGATTAT CCTAGTGCCA GAGGTGGTTT AGAAGATATC 2640  
CTCATCATGG AAAATATGAC CAATCATCTC CTTTGGTTT AAATCCGAGT GCATGGCTAT 2700  
TTGCTTGATT TTGCTAGTAT TGAAGGGCAA AGGCAAAAGC ATTATCGTTT GAAAAATTTA 2760  
CCTCAGACGG TTGAACTGAC AGTGGATGAT GTGGAGGAGG ATGTGGATTT GACCCTACCT 2820  
GAAAAATCGAA GTTATCAAGA AGCTGATTTT TTTGAACGCA TGTTCGAGA GAACTGCTAA 2880  
GGCCACTTTT AAAGATTTC AAGACTATCT TTCTTCATGA GGAAAGATAG TTTTGTGGTA 2940  
TGATTTTCAT TCCCAAAATA CAAGGGGAAT GTGTTACAAT AGTAGTAACA GATAATAGAA 3000  
AAGAGAATAG ATGAGAATTG CAGATTATAG CGTGACCAAG GCAGTGCTGG AGCGTCACGG 3060  
TTTTACCTTT AAAAAGTCCT TTGGGCAAAA TTTTGTGACG GATACCAATA TCCTTCAAAA 3120  
AATTGTGGAT ACGGCTGAAA TTGATGATCA GGTCAATGTC ATCGAAATCG GGCCAGGTAT 3180  
TGGTGCCCTG ACAGAATTTT TGGCTGAGCG TGCAGCCCAA GTCATGGCTT TTGAGATTGA 3240  
CCACCGTTTG GTGCCAATTT TGGCAGATAC CCTGCGTGAT TTTGATAATG TGACCGTAGT 3300  
TAACGAAGAT ATTCTCAAGG TTGATTTGGC GCAACATATC CAGAATTTTA AAAATCCTGA 3360  
CCTGCCAATC AAGGTAGTGG CTAATTTGCC TTAATACATC ACGACGCCTA TTCTCATGCA 3420  
CTTGATTGAG AGTGGCATTC CTTTTGTGA GTTTGTGGTC ATGATGCAGA AAGAAGTAGC 3480  
GGACCGCATT TCAGCCCAGC CTAACACCAA GGCTTACGGT AGCTTGTCTA TCGCCGTGCA 3540  
GTATTACATG ACAGCCAAGG TTGCCTTTAT CGTGCCTCGT ACGGTCTTTG TGCCAGCGCC 3600  
AAATGTGGAT TCAGCCATCT TGAATATGGT GCGTCGTCCA GAGCCAGCCG TAGCAGTAGA 3660



1001

AGATGAGAAC TTTTCTTTA AGGTTTCCAA GGCTAGTTTT ACCCATCGCC GCAAGACCTT	3720
GTGGAATAAC TTGACAGGTT ACTTTGGTAA GACTGAAGAG GTCAAGGACA AGCTGACCAA	3780
GGCTTTGGAC CAGGCAGGCT TGTCACCAAG TGTGCGTGGG GAAGCTCTCA GCTTGGCAGA	3840
ATTTGCCGGT CTAGCAGACG CACTTAAAGG GCAAGGACTC TAAGATGCAG GGACAAATCA	3900
TTAAAGCCTT GGCAGGTTTC TACTATGTGG AGAGTGATGG CCAGGTTTAT CAAACACGCG	3960
<hr/>	
CGCGTGGGAA TTTCGGTAAA AAAGGCCATA CCCCTTATGT TGGGGACTGG GTAGATTTCT	4020
CTGCCGAGGA AAATTCAGAA GGCTATATCC TCAAAATTCA CGAACGGAAA AACAGTCTGG	4080
TTCGTCCGCC TATGTCAAT ATCGATCAAG CTGTAGTAAT CATGTCCGTC AAGGAACCTG	4140
ATTTAACAG CAATTTGCTG GATCGTTTCT TGGTTCTTTT GGAGCACAAG GGCATCCATC	4200
CCATTGTCTA TATTTCCAAA ATGGATTTGT TGAAGATAG GGGAGAACTG GATTTTACC	4260
AGCAGACCTA TGGTGACATC GGCTATGACT TTGTGACCAG TAAAGAGGAA CTCCTGTCTT	4320
TGTTAACAGG CAAGGTTACG GTCTTTATGG GGCAGACAGG TGTGGGAAG TCAACTCTTC	4380
TCAATAAAAT CGCACCAGAC CTCAATCTTG AAACGGGAGA AATTCAGAC AGTCTAGGTC	4440
GCGGTCGCCA TACCACTCGA GCTGTTAGTT TTTACAATCT CAACGGGGGT AAAATCGCAG	4500
ATACACCAGG ATTTTCATCC TTGGACTATG AAGTATCAAG GGCTGAAGAC CTCAATCAGG	4560
CTTTCCAGA GATTGCTACT GTTAGCCGAG ATTGTAAGTT CCGTACTTGT ACCCATACCC	4620
ATGAGCCGTC TTGTGCCGTC AAACCAGCTG TTGAAGAGGG TGTATTGCA ACCTCCGTT	4680
TTGACAATTA CCTGCAATTC CTTAGTGAAA TTGAAAATCG TAGAGAAACC TATAAAAAAG	4740
TCAGCAAAAA AATCCAAAA TAAGGAGAAA CCTATGTCTC AATACAAGAT TGCTCCGTCA	4800
ATTCTGGCAG CAGATTATGC CAACTTTGAA CGTGAAATCA AACGTCTAGA AGCAACTGGG	4860
GCAGAAATG CCCATATCGA TATCATGGAC AGTCATTTTG TACCGCAAAT CAGTTTGGT	4920
GCAGGTGTGG TCGAGAGCCT TCGTCCTCAT AGTAAGATGG TTTTCGATTG CCACTTGATG	4980
GTGTCAAACC CTGAGCATCA TCTGGAAGAT TTGCGCGTG CAGGTGCAGA CATCATCAGT	5040
ATCCATGTAG AAGCAACGCC TCATATTCAT GCGGCCCTCC AAAAAATTCG TTTACTCGGA	5100
GTTAAGCCTT CAGTCGTTAT CAATCCTGGC ACATCAGTTG AAGCCATCAA GCACGTCCTT	5160
CATCTAGTTG ACCAAGTTTT AGTCATGACG GTTAATCCAG GTTTTGGTGG GCAAGCCTTT	5220
CTGCCAGAAA CCATGGATAA GGTCCGTGAG TTGGTTGCTC TTCGTGAGGA AAAAGGTTTG	5280
AACTTTGAAA TCGAAGTGA TGGTGGGATT GATGACCAAA CTATTGCTCA AGCCAAAGAA	5340
GCCGGTGCGA CTGTTTTTGT AGCAGGTTCC TATGCTTTA AGGGAGAACT CAATGAGCGA	5400

1002

GTACAACTC TCAGAAAACA ACTGGACTAG GGTTCAGTT TTTGCAGGCG GAAACCGCGG	5460
TCATTATCGG ACAGATTTTG ATGCTTTTGT TGGGGTGGAT CGAGGCTCGC TCTGGGTCTT	5520
GGAAGAAGAC TTACCTCTTG CTCTAGCAGT CGGAGATTTT GATTCTGTGA CGGAAGAAGA	5580
GCGACAGGTG ATTCAAAAAG GTGCCCAGTA TTTTGTCCAA GCACGACCAG AAAAGGATGA	5640
TACAGATCTG GAATTGGCTC TCTTAACCAT CTTTGAACAA AATCCTCAGG CTCAGGTCAC	5700
TATTTTCGGT GCCTTGGGTG GCCGTATTGA CCATATGTTG GCCAATGTCT TTCTGCCTAG	5760
CAATCCTAAG TTGGCACCCT ATATGCATCA AATAGAAATT GAGGATGGGC AAAACTTGAT	5820
TACTTATTGT CCAGAAGGAA TCAGTCAGCT AGAACCTCGT TCAGACTACG ACTATCTAGC	5880
CTTTATGCCA GTTCGGGATA GCCAGCTGAC TATTCTTGA GCCAAGTATG AGTTGACAGA	5940
GGAAATTTT TTCTTTAAAA AAGTGACGC TTCTAACGAA TATATAGATA GGGAAGTGTC	6000
GGTAACTTGC CCAGATGGTT ATGTGGTCGT ACTGCATAGC AAGGACAGGA GGTAGGATGG	6060
AAAGTTTACT TATTCTATTA TTAATTGCCA ATCTAGCTGG TCTCTTCTG ATTTGGCAAA	6120
GGCAGGATAG GCAGGAGAAA CACTTAAGTA AGAGCTTGA GGATCAGCA GATCATTGT	6180
CAGACCAGTT GGATTACCGC TTTGACCAAG CCAGACAAGC CAGCCAGTTA GACCAAAAAG	6240
ATTTGGAAGT GGTGTGCAGC GACCGTTTGC AAGAAGTGC GATTGAATTG CACCAAGGTC	6300
TGACCCAAGT CCGTCAAGAA ATGACAGATA ATCTCCTCCA AACTAGAGAC AAGACAGACC	6360
AACGTCTCCA AGCCTTGCAG GAATCAAATG AGCAACGTTT GGAACAAATG CGCCAGACGG	6420
TCGAGGAAAA ACTAGAAAAG ACCTTGCAGA CACGCTTACA GGCTTCCTTT GAGACAGTTT	6480
CTAAACAAC TGGAGTCTGTC AATCGTGGCC TTGGAGAAAT GCAGACAGTT GCCCGTGATG	6540
TCGGAGCTCT TAACAAGGTT CTCTCTGGAA CCAAGACGCG AGGGATTCTG GGAGAATTGC	6600
AAC TGGGGCA AATTATTGAA GACATCATGA CACCTGCCCC GTACGAACGA GAATACGCAA	6660
CGGTTGAAAA CTCTAGTGAA CGAGTGGAGT ATGCCATCAA GTTACCCGGA CAAGGCGACC	6720
AAGAATACGT CTATCTGCCA ATTGACTCTA AGTTTCCACT GGCAGATTAT TACCGCTTUG	6780
AAGAAGCCTA TGAGACAGGT GACAAGGATG AGATTGAACG CTGTCGTAAG TCACTCCTAG	6840
CAAGCGTCAA GCGCTTTGCT AGGGATATTA GGAACAAGTA CATAGCACCA CCTCGGACGA	6900
CCAATTTTGG AGTTTGTGTT GTTCCGACAG AAGGTCTCTA CTCAGAAATC GTCCGCAATC	6960
CGGTCTTCTT TGATGATTTG AGACGGGAAG AACAGATTAT TGTTCAGGA CCAAGTACCC	7020
TATCAGCCCT TCTTAACTCC CTATCAGTTG GTTTCAAGAC CCTTAATATC CAAAAGAGTG	7080
CCGACCATAT CAGCAAGACT CTTGCCAGTG TCAAGACCGA GTTTGGCAAG TTTGGTGGTA	7140
TTCTGGTCAA GGCACAAAAA CATCTCCAAC ATGCCTCTGG CAATATTGAT GAATTATTAA	7200

1003

ACCGTCGTAC CATAGCTATC GAGCGGACGC TCCGTCACAT TGAGTTGTCA GAAGGTGAGC	7260
CTGCCGCTTGA TCTACTCCAT TTTCAAGAAA ATGAGGAAGA ATATGAAGAT TAGTCACATG	7320
AAAAAAGATG AGTTATTTGA AGGCTTTTAC CTAATCAAAT CAGCTGACCT GAGGCAAAC	7380
CGAGCTGGGA AAAACTACCT AGCCTTTACC TTCCAAGATG ATAGTGGCGA GATTGATGGG	7440
AAGCTCTGGG ATGCCCAACC TCATAACATT GAGGCCTTTA CCGCAGGTAA GGTGTGCCAC	7500
ATGAAAGGAC GCCGAGAAGT TTATAACAAT ACCCCTCAAG TCAATCAAAT TACTCTCCGC	7560
CTGCCTCAAG CTGGTGAACC CAATGACCCA GCTGATTTC AAGTCAAGTC ACCAGTTGAT	7620
GTCAAGGAAA TTCGTGACTA CATGTCGCA ATGATTTTCA AAATTGAAAA TCCTGTCTGG	7680
CAACGGATTG TCCGAAATCT CTACACCAAG TATGATAAGG AATTCTACTC CTATCCAGCT	7740
GCCAAGACCA ACCACCATGC CTTTGAAACG GGCTTGGCCT ATCATACGGC GACCATGGTG	7800
CGTTTGGCAG ACGCTATTAG CGAAGTTTAT CCTCAGCTCA ATAAGAGCCT GCTCTATGCG	7860
GGGATTATGT TGCATGACTT AGCTAAGGTC ATCGAGTTGA CGGGGCCAGA CCAGACAGAG	7920
TACACAGTGC GAGGTAATCT TCTTGGACAT ATCGCTCTCA TTGATAGCGA AATTACCAAG	7980
ACAGTTATGG AACTCGGCAT CGATGATACC AAGGAAGAAG TCGTTTGTCT TCGTCATGTC	8040
ATCCTCAGTC ACCACGGCTT GCTTGAGTAT GGAAGCCCAG TCCGTCCACG CATTATGGAA	8100
GCAGAGATTA TCCATATGAT TGACAATCTG GATGCAAGCA TGATGATGAT GTCAACAGCT	8160
CTTGCTTTGG TGGATAAAGG AGAGATGACC AATAAAATCT TCGCTATGGA TAATCGTTCC	8220
TTCTATAAAC CAGATTTAGA TTAATAATTT AAGAAAAATG AGCATTTTTT AGGATAAGAA	8280
TGTTCCGTTTT TTTATGTGAA TATGGTATAA TAAGTAAAAG ACAAAAATGA ATACTCTTCG	8340
AAAATCTCTT CAAACTAGGG TAGTATCGCC TTGTCGTATG TATATATGCA GGTATATTAC	8400
AGGGTTTGTC AGTTCTATTG ACAATCTCAA AACAGTGTTT TGAACCACCA GCGACCAGCT	8460
TTCTAGTTTG CTTTTTGATT TTTTGAATAA AAATGGAATA GGAAATAGAA ATGAAATTAA	8520
GAAGAAGTGA TCGGATGGTT GTCATTTCCA ACTATTTGAT TAATAATCCT TATAAACTAA	8580
CTAGTCTCAA TACTTTTGCT GAAAAGTATG AGTCTGCTAA ATCATCCATC TCAGAAGATA	8640
TCGTCAATTAT CAAACGCGCC TTTGAGGAAA TTGAAATCGG TCATATCCAG ACAGTGA	8700
GGGCTGGCGG AGGTGTCATC TTCACACCGT CTATTTGAG TCAGGATGCT AAGGAAATGG	8760
TTGAAGACTT GCGTACCAAG TTGTCAGAAA GTGACCGTAT CTTGCCAGGT GGTATATCT	8820
ATCTGTCTGA TTTGCTTAGC ACACCAGCCA TCTTGAAAAA TATTGGTCGT ATTATTGCCA	8880
AAAGCTTTAT GGACCAAAAA ATTGACGCGG TTATGACCGT AGCAACTAAG GGTGTGCCAC	8940

1004

TTGCAAATGC	AGTTGCCAAT	GTCCTCAATG	TCTCTTTTGT	CATTGTGCGC	CGTGACCTGA	9000
AAATTACCGA	AGGTTCAACT	GTTAGCGTCA	ACTATGTTTC	AGGTTCAAGT	GGTGACCGTA	9060
TCGAGAAAAT	GTTCTTTTCA	AAACGTAGTC	TTAAGGCAGG	CAGCCGTGTC	TTGATTGTGG	9120
ATGACTTCTT	GAAAGGTGGC	GGAACGGTCA	ATGGTATGAT	TAGTCTCTTG	CGCGAGTTCC	9180
ACTCAGAACT	GGCAGGTGTA	GCGGTCTTTG	CGGACAATGC	CCAAGAAGAA	CGTGAAAAGC	9240
AGTTTGACTA	CAAGTCACTC	TTGAAGGTAA	CCAATATTGA	TGTCAAGAAC	CAAGCCATCG	9300
ATGTTGAGGT	TGGCAATATC	TTTGACGAAG	ATAAATAAGA	GATAGAACTA	AAGGTTGGAA	9360
CGATTGTCCC	AGCCTTTCTT	TGCAAACAGA	ATAGAAGGAA	GCTTATGAAA	ACACCATTTA	9420
TCAATAGAGA	AGAGTTAGAA	GCGATTGTTG	CCGAGTTCCC	GACTCCCTTT	CAC TTGTATG	9480
ATGAGAAGGG	GATTCGTGAG	AAGGCAAGAG	CCGTCAACCA	AGCTTTTTCG	TGGAACAAGG	9540
GCTTTAAGGA	ATATTTTGCA	GTTAAGGCTA	CTCCAACCTC	AGCTATTTTG	AAAATTCTCC	9600
AAGAAGAAGG	TTGTGGTGTG	GACTGTCTCTA	GTTATGTAGA	GCTTTTGATG	AGCCATAAAC	9660
TGGACTTTCT	GGGTTCTGAG	ATTATGTTCT	CTTCCAACAA	CACGCCAGAC	AAGGAATACG	9720
CCTATGCACG	TGAATGGGT	GCGACCATTA	ACTTGGATGC	CTTTGAAGAT	ATTGAACATC	9780
TGGAGAGAGT	AGCAGGCATT	CCAGAAATCA	TCTCTTGTCG	TTATAATCCT	GGAGGCGTTT	9840
TTGAACTGGG	GACAGACATT	ATGGACAATC	CTGGGGAGGC	TAAGTTTGGC	ATGACCAAGG	9900
ACCAGCTCTT	TGAAGCCTTT	GCTATCTTGA	AGGAAAAAGG	AGCCAAGACT	TTTGGGATTC	9960
ACTCCTTCCT	AGCGTCCAAT	ACCGTGACCC	ATCTCTATTA	TCCAGAGTTG	GCTCGTCAGC	10020
TCTTTGAACT	GGCTGTTGAA	ATCAAGGAAA	AGTTGGGCAT	TTGCTAGAC	TTTATCAATC	10080
TTTCTGGCGG	TATTGGTGTT	AATTATCATC	CAGACCAGGA	GCCGAACGAT	ATCGCCTTGA	10140
TTGGTGAGGG	AGTTCGTAAG	GTGTATGAAG	AGGTTCTTAC	GTCAGCAGGT	CTTGGTCAGG	10200
TCAAGATTTT	CACCGAATTG	GGTCGTTTTA	TGCTGGCACC	TCACGGTGCT	CTAGTCACAA	10260
GAGTCACTCA	TAAGAAGGAA	ACCTACCGTA	CCTATCTAGG	TGTGGATGCC	TCAGCAGTCA	10320
ACCTCATGCG	TCCAGCTATG	TACGGAGCTT	ACCATCATAT	TAGCAACGTG	ACCCATCCAG	10380
ATGGACCAGC	TGAAGTGGTA	GATGTGGTCG	GTTCACTCTG	TGAAAACAAT	GATAAATTTG	10440
CAGTTAATCG	CGAACTGCCT	CATACAGAAA	TGGTGATTTT	GCTGGTCATT	CATGATACAG	10500
GTGCCACCGG	ATTTTCAATG	GGCTACCAGT	ATAATGCCAA	ATTACGTTCT	GCGGAAATCC	10560
TCTATACCGA	AGAAGGTAAA	GCCCGTCAAA	TCCGCCGTGC	AGAGCGCCCT	GAGGACTATT	10620
TTGCAACCTT	ATATGGCTTC	GATTTTGAAG	AATAATCTGA	TAATAGATTG	AAAATGAAAT	10680
TGAAAAACAG	ATTGCTTTCT	AAAAAATAGG	CAAAAATCTT	GTTTTTCCTT	CAAGTCGTGA	10740

1005

TATAATAAAA CTATAAACG TTTTCAAGGA AGGTAACGAT ATGTCTGAAG AAACAATTGA	10800
TTATGGACAA GTGACAGGAA TGGTGCATTC GACAGAAAGC TTTGGGTCAG TAGATGGGCC	10860
TGGTATTCGC TTTATTGTCT TTTTGCAGGG CTGTACATG CGTTGCCAGT ATTGCCACAA	10920
CCCAGACACT TGGGCTATGG AGTCCAATAA GTCACGTGAA CGGACGGTAG ATGATGTCTT	10980
GACAGAGGCC TTGCGCTACC GTGGTTTCTG GGGAAATAAG GGTGGGATTA CAGTCAGTGG	11040
AGGAGAAGCT CTCTTGCAGA TTGATTTCTT GATTGCTCTC TTCACCAAGG CTAAGGAACA	11100
AGGAATCCAC TGTACCTTGG ACACCTGTGC TCTTCCTTTC CGTAATAAAC CACGTTACCT	11160
TGAGAAGTTT GACAACTCA TGGCTGTCAC TGACTTGGTT CTTTGGATA TCAAGGAAAT	11220
CAACGAAGAA CAGCACAAGA TTGTCACTAG CCAAACCAAT AAAAAATCT TGGCTTGTGC	11280
CCAGTATCTA TCAGATATTG GAAAACCTGT CTGGATTCCG CACGTGCTAG TTCCAGGATT	11340
GACAGACAGA GATGATGACT TGATTGAACT TGGTAAGTTC GTCAAGACCC TCAAAAATGT	11400
TGATAAGTTT GAAATTCTAC CTTATCACAC CATGGGTGAG TTCAAGTGGC GTGAACTTGG	11460
AATTCCATAT TCCCTCGAAG GAGTCAAACC ACCAACAGCA GATCGCGTCA AGAACGCTAA	11520
ACAACTCATG GATACCGAAA GTTATCAAGA TTATATGAAA CGTGTACATG GATAGAAAAG	11580
AAGCCTGATG GAAACATCGG GCTTTTGA CTGAAAAAGA CTTAGCAAAT CAGCTAAGCC	11640
TTTTTCTTCT TATCTCGAAC GTTGTTTTCC AGCGTTGCGA TTTTGTGTT TTTCTTGCT	11700
TGTGATAGCA GTTGGTTGTT CAGGGGTAAC GTCTTTTCGT CCACTTGCTT TAGAGAAAGC	11760
ACTTGCTTTT GGTGGGTCTT TGGCTAGTTC TTCACGGACT TTTTGGCGAA GTTTGGACG	11820
AACGATATAG TTGACGATAA ACTGTTGGAG AATCATCATG AAACCACCGA CAACCCAGTA	11880
AAGTGTGACA CTAGCTGGTG AGAAGAGGGA GAAGACGACG ATCATGAGTG GGCTCATGTA	11940
AATCATTTTC TTGATTTGTT CTCTTTCAT TTCATCTTCT ACTCCGTGAA GTGAAAGGAG	12000
CGATTGAAGA TAGTAAAGGA CACCAGCACA GGCAACCAAA ATCATACTTG GAGAACCCTAG	12060
AGGAATGCCT AGGTAGCTTG CTTGAGCAAC CCCTTCAGTA TGTTGGGCAG CAAAGTAGAT	12120
AGCAGAGAAG AAAGGCATTT GAAGGAGGAT AGGGAAACAT CCTACACCGC CAAACATGCT	12180
GATACCGTGC TCTTTTGTAG CAGCAAAGAG AGCTTGTTGG GCTTCGAGTT TTTCTTCTTG	12240
AGTAGTCGCT TCTTTGAGAC GCGTTTGGTG TGGCTCAAGG ACGTGCTTGA GGGCGTTCAT	12300
CTTTTCAGAG TGAAGCGTTG CCTTCCATGA TTGGTAGATA CCAAGTGGTA AGATAATCAA	12360
GCGTACGATA ATGTTTACGA TAATGATAGC GACACCAAAG CCTAGACCTT TATCAGTAGC	12420
GAAGTACTTG ATGGCTTCAG CCATAGGCGC TCCGATCGTA TTCCAAATAA ATCCTGTTGG	12480

1006

CTGACCTGTG GTTTTATCGA CATTGACACA GCCAGTCAAG ACAAGCAACA TAGCCACTCC	12540
CATAGCCGAG AGTGCAAAAT CGGGGT	12566

## (2) INFORMATION FOR SEQ ID NO: 150:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5238 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

TGACACTCTG TAGGATTGTC GTTAATTGAT TGCTCGTACT CTCTACAATA ACCACCAAAG	60
TAAAAACGAC ATAGAAAGAT AGCATCAGCT GTAGCCATAG CGCCTTTGAC ACCTTCTGGA	120
TGATTATGAG TTACCTCTGC AGAAAGACTC GTAAGTCCTC TAGATGATGG CCATATACCA	180
GTTTTCGCAT AAAAACCACA GTCCATGATC CAAGCACATG GAGAAATACG CATAGCTGAT	240
CCATTCCCAA AGCTATTATA AGGCTCACGG TTATCGCTGT TTAGCCATGC ATTAAACCGA	300
GCACCGTAAT CAGCATTCGG ATACATTCTG CCATATTTCT TCATCGCGTC AATGAAGTCA	360
TCTTTTGTG CACCATTTCAT AATTGCTTCT GCAACAGCAC AGGTCATAAC CGTGTCATCT	420
GTAATAAAGC AGTCCTTCCG AAATAAAGGA AAGTCCTTTG TTTTGATATT GTTCCATTCC	480
TAAACAGAAC CGACAATATC TCCAATAATT GCTCCAAGCA TCAGATTCCT CTTGTTCAT	540
TTTGATGCTT TTTATATTGG TTATCTACCA TATTTATTTT AGAAAATAAC ATCCTGTTGG	600
ATTTTAAAAA TTTTATTTT TTCAAAATAG GGTTTACCA TTTCTTTCCA CCTAGCTCTA	660
TGAAAATTGA TTGATTTTAA AGGAGATAGG CCATAATTTT CCAATGCATA ACCATCATTT	720
ACTTCAACAA CAAGTGTCTT GCCATCGCGA GTAACACCGA TATCTAGTCC ATAAGCTATT	780
GGCGCATCTT TCCAACATGA TATCGCTTCA TCAATTACAC TTGCATCAA TTGTGCATGA	840
TAATCACCTG TATAGGTCG AACATCTAAT ACGCGACCAT CTAACACAAA ACAACGCCAT	900
TCAGCTATGA ATTCTACAAC CTCCTAATC CATATAGGAT AGTCGAAAGG TAGACCAATA	960
CCTATTAAAT CATGGGTCC ATTAACAACT CTCCAGTAA AGACTTTTGA ACCAGCTTTA	1020
GGCTTAATAA ATTTTCCCCA ATTATCAGGT ATATTACAA TCTCTCTAA AATACCAGCA	1080
TAAATCTTTC GACCATAAAA CTCTTTAAGC TCAATAGGAT AGTCATGAAC CGGAACGTTT	1140
AAGCCCATCA TTTTATGTA TGCTCTAGTC TCCATTATAT AATCTACAAC TATATCTTCA	1200
CTTGTTAACT CTTTATTTT AGAAAAAGAT TGATATAAAA TAACTTCTTC TCCTTGTAAG	1260
TAGGCACCTA CTTGAGCATT GTATTTATTA ATTGAAACCT CACTTGGTAA TTTACTTTGT	1320

1007

CTAATATAAA CAACCATTTC ATCACTCCTA TATCACTAGT GTTACACCAA TTTGTAAAAA	1380
ATAATAGCAA TTTTGCTCTT ATTTTTTTGA GTAAATAGCC CCCATAATAT CATCGAAATA	1440
ATCAACGGTA TTTAGGAGTA ATTCAATAAC CTGGGACTTT GTTAGTCGCA TTCCCTTCT	1500
ATCTCTAGCA TCTTCTACTA AATTTTCAAG TTTCTCTAGA TTTTATCAT CCAAGCTAAT	1560
CATTATTCTA TTTTATCGG TTGCCATTTT CATCACCTCA AGTTAATTCT ATCACAGGTG	1620
TAACACTAGT GTCAACTGGC TTTTATAATA CATTAGTTTA AAAGTGGAGA GGATTTTAA	1680
CACAGTAACT TTAAATCTTT GGTATTAAAA AATTTTCACA ATATTTATAG AAATAAAATC	1740
TGTCTCAAAT CAGTTATCAA ATCTAGTATA AATTATGAGC GGCTACTCTA ATACTTTCCC	1800
TCTAAACAAG AAAAAGACTT ACACTCAAGG GTTTTCTTCC CCCCCTTCGT TATAACGTTT	1860
TGACTCTTTT ACTAGCAAAG GTATATACTC ACAAGGAACT TTGGTTGACT ATTGAATCTC	1920
TCCAACCTCT TCTTTAACAT ATCCTTCTAC ATCTTCAATC TCTACAAACA TTGGGTCTAA	1980
GTGACACAAG AAATGCCAAA CTTCGATCCC TTTTTTCTG TAAAGAATCG CTTCACCGTC	2040
TTCACTTCCG AAAAAGCTTC TGTCGATTTC ATATCCGCGG CTTTCTAAGA AGTCTTTTGC	2100
TTTACGATAG TTCGTTTCTC TTGTTTCGAC ATAGGCTTTA ACTTCATGGT TGTAAACGAC	2160
ATATGCATCA ATTTTGAAT ATCCTTCGAT CACTCTATCA TTTTGAGGG ATAAATTGTA	2220
AATCTCTTTC CAAATAATGT TTACATTTTC CTCAGGATCG AACATAAATT TAGATAAAGG	2280
AACAATATTT CCGTTAAAAA TAATTTCCAT ATAATCCGGT ATGTTTTTAG GATTAATAA	2340
CTCCACTTCA AAACCATCTT CTGTTTCCAG AGTGTATCCC GGGATTTGAG CTACAAAGGC	2400
TTTCCCATCT TCTATGGAAT CAAATGCTAC TAAATCTTTA GAATAATCAT TTTGGTACAA	2460
TTCCAATATA ACCATCGATA ATCTCTCCAT TTTCAATATC AGGCTAATGT AAATAAGCAC	2520
GTCACCTGAC CAATTCAGGC TCTCTGTATC ATCTCATCAT ATTCCTACT TACTTTACGA	2580
GTCTTATACC CAGAACACAC CTTATCGACC TTCGGTCTCA CCTCGTCGCA TTGGCTGAAC	2640
ATCTACTTTT ACTTTGCTGA TGCTTCAACT CGTACAAGCA GTGATACCGC CTCAGCGTGA	2700
TGCGTCAGTG GGACTCAAAA GGTTCCGGGA ACCTTTTGAG GATTAACTAC GTTCTCTAA	2760
TAACTTACA CATTCAACTT GTTCATCAT GTCCAAACCT ATGTTGAGAT TTTCTTCTAT	2820
AATTGGTAGC TTAAAAGTAA TGGATTTTAG CCATTGTCCG TTAGATTGTT TTTCTTCATA	2880
AACTTGAATT TCAGAAATCA AAGCTGAAAT TAACTGCCTA CGCTCTACAT CATTATGAC	2940
TTTATAGAGC TTATCAAAAT AGATCAGAAC CTTATATATG TTATCTCCTG TAAGCTTTTC	3000
AGCTTCAATA GTCTGTTTCT TTGCTTTCGC ATCAATTAGT GATGATTCTA ATTCATCTAG	3060

1008

TTTGTACATAC ATACGATATA GTCTATCATC TAAATCCTGT TTCCTTCTCT TATAATGCTT	3120
ATCTTCAACA TCTAAATTAT CTATTTCTCTC AATTAGCTTA AACTTTGTAG AATGACTCTT	3180
TCTCAATTCC TTTTGGTAAT TATCTATTTT TTTTCTATT TCAGAGGTAT CCACCTTCAT	3240
GTTGATTTTT TCTTGCATCA TAGAAGCAAA TTTCGGATTA CTTACTATCT TGACAATCAC	3300
CTCTGCAACA GCATCATCTA ACAATTCTTC TCTAATTTGC TTACTGAATG TACACTTATT	3360
ACCTCTTATC ATCTGCCTAT GGTTACAACC ATAGTAATAA AAATCTTTAT ACTTTGTGCC	3420
ATCTTTCTTT TTCTTGATAC ACTTGTTCCT AAACATTCCC ACTCCACATA TCGGGCATT	3480
TACAATTCCA GAAAGCAAGT GTGTGCGTGT ATCTTTTCTT TTATTCACAT GCTCATATTT	3540
CTTTGCTTGA GATTTTAGCT TAACCTGAGC AGCTTGCCAA ACTTCATCGG AAATATAGC	3600
TTCATGTATC CCTTCAGATA TTAGATATTC ATCTTGTTCA ACCTGCTTAT ATTCATTTCT	3660
TGTACCATGA ACTTTTTCTA AAGTTCTTCT TCCAAATGCT ATTTTCCCAT TATATACAGG	3720
ATTCTTTAAT ATCTTTCTTA TAAGACCTGC ATCAAACAAA GGATTCTTAC CATTCTGTCT	3780
TGGGATTTTT CTAATTCCAT GATTCTCTAA GTATTTAGAT ATCCCATTGG CTCCTATCGT	3840
AGTATTTACA TACTGGTCCA AAATCGTTCT TATTGCAACT GCCTCTTCTT CATTTATAAA	3900
CAGCTTGCCG TCTTCAAGTT TATATCCATA CGGAGCAAAG CCACCATTCC ATTTTCTTCT	3960
CCCTGCTTTT TGAATGCGAC CTTCCATTGT TTGAATACTG ATGTTTTCTC TTTCTATTTT	4020
AGCCACAGCT GATAAACAG AAATCATTAG TTCCCAGCA TCTTTAGATG AATCAATGCC	4080
ATCTTCAACG CAGATAAGAT TAACTCCATA ATCCTGCATT ATATGAAGTG TAGAAAGAAC	4140
ATCAGCGGCA TTTCTTGCAA ATCTTGATAA CTAAACACA AGAACAAAAG ATACTCCATC	4200
TTTCCAGAT TTTATATCTT CCATCATTCG ATTGAACTGT ATTCTACCTT CAATAGACTT	4260
GTCAGACTTC CCGGCATCTT CATACTCTCC AACAATTICA TAATCGTTGT AAATAGCAAA	4320
AGCTTTCATT CGTGATTTTT GTGCCTCTAA CGAATACCCC TCTATCTGTA TTGACGTAGA	4380
TACTCGTGTA TAGAGGTATA CTTTTATTTT TTCTTTTGAC ATAGTATTAA CCTCAATATA	4440
ATTTTCTAT ATCATATATA ATTTTPTAA TTAAAGTTTG GACTATCATT TCAAGTATAT	4500
TATAACACTT TTATTAGTCC GTCTCAATTT GTGTTTTTGC CATGTCAAAA CTATTTTCA	4560
TCTCTTGATT TTTTGCTGGC GTTGGATCGG GTAGATTATC TAAATCTAAA GCACCAGCAT	4620
ATTTTGCAAT CAGATTGCT ATTAATCAG CCAATCCATT CCAGTCATTG TCCAATATAT	4680
ACCTCCTCTA AAGTTTTATA TCTAATAATT ATTTGTTTAA TTAAGTTTTT TGACATTGAC	4740
AAGTGCTTTG GATTAGCAAC ATAGGAATCT CACTTCCGCC TCTATTCCGG ATGAGCCGGC	4800
TTCAACCTTA GAAGTATCAT TACCCTCATT TTCTTCATAG CGGATAGGGT ATCCCTCCTT	4860



1009

ATATTCAAAC TCTTACTTAT CGCTCACTTT CTTTTTGCTT AGCAGAACTT TTTTGGCCGA	4920
ATTATTACAGC CGAAAGATCT TGACGGATAG GTTATTACGC TCCAAAAATA ATTAACGTCT	4980
TGTCTTGGTC TATTCAATTG TTAAGGTTCA AAATTTATCG AGAGTTATTA ATCTTTTAA	5040
AATTTGACCA TCAGAAAATA TTTATCTTGA TGTAACAAAA TTCTATAAAT TACCCTCTTA	5100
TACTTAACAG TGAAGAAG TCTTCTTGG TAACCAATTT TGAAATAGAA TTTGCTTATA	5160
TAAAAAGGTC CAATTCCAC TGCATAAATA GCAGTAAAA TTAGACCCTC TTGGTAACTG	5220
TCATCTAAAA GTCTTCTA	5238

(2) INFORMATION FOR SEQ ID NO: 151:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 13425 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

GACGATTTAC GAAGAATCGA ACAAGAACCT GCTCCTATCA ATTCCCAACC TCTATCTCTA	60
AAATCTTGCA GTTCATGCTT ATACTTTTTT AAGAAATCTA GAATCATAGA TACGGTAGAT	120
GACATCGTCT GGTGACATT GGTCAAAATA GAACAAACCA AAACGACTCG TTCTATACCT	180
CCAACCTTTC AAATGCATCT CATGTAAATG TTCTTCTTCC TTGTCCAAAT CAACAATGGT	240
GAAATCCGA AATCTACTC TGCTATTCAT TGTCTTACCC CAAAATTAGA AAACATGCCT	300
GGCGTTATTT ATTAGATAAT TCTTCCACT TTTGACTCAA TCTCCAAAA ATATAAGAAA	360
TCTGAATCGC AAAAATATC AATAAAACCC AATCTATTAT GAAAATCAA AACACTTTCC	420
AACTGAAAGA ACTACCTCCA GTGACAACT TTGAGAAAA CGGTAGTAGA GCTAAAAAGA	480
GAAATAAAT AGGAAGCATC CGCATTGTTA AAATCCGTTT GGCATAAAAA AATCTTTATT	540
TAAACGAAAA TATTATGGCA AAATTTACGC CAGTTTTTGA ACGGCTGATG TAGATATTTT	600
ATACTTTCAA AATGTTTAAA TGTGATTATT TATTTTGA AAATAGATCA CCAGCCCGAC	660
TGAAAGTGCT TATAGAATGA TAATAAGTCG CCTGCCGAAA ACAGCGAAAA ATAGCGGTGT	720
TATGCGGAGA TAATCTGACG CGATGCGAAA GTATATTGCA TACTTATTTT CAACAATTTA	780
GCAGAGTATT TTTATAAGTG TGATATAATA GAAGTATAAT TTGTTCTGAT AGTTTATTTT	840
ATGGAGAACT AGATTTTTAG AATGCGGAGG GTTCAATATG GTTGAGTTTA TAAAGTCTAA	900
GAAAGAAATG AGTGAGGAGG ATATTAAAGC AAATTCATC ACTCCTGCTA TTGTATCCAA	960

1010

AGGATGGAAA AATGGTGAGC ATATCGCTTA CGAAGAATAC TTCACTGATG GTCGAATTGA	1020
AGTTAGAGGA GATAAGGCTC GTCGTAAAGA AGGAAAAAAA TCAGACTATT CACTGTATTA	1080
CCAATTTGGA ACTCGAATTG CAATTGTTGA GGCAAAGGAT AATAAACACA GCGTTCGAGC	1140
AGGATTACAA CAAGCTATTG AATATGGAGA GATTTTAGAT GTTCCATTG TTTATTCTTC	1200
GAATGGTGAT GGCTTTATTG AACACGACCG TATCACGAGA GAAGAACGTG AGCTGGAGTT	1260
AGACGAATTC CCTACTCGTG AAGAATTATT TTCTCGTATG ACGAAGGAAA AAGGATTGAC	1320
GTACGAAATT ACAGAAGCTA TCTCAACTCC ATACTATACA GACGCCTTCT CAATGAAAAC	1380
GCCACGCTAT TATCAGCAAA TAGCTATCAA CCGTACTATT GAAACAGTTG CCAGAGGACA	1440
AAAACGAGTA ATGTTTGTGA TGGCAACAGG AACGGGAAAA ACGTTCATGG CTTTTCAAAT	1500
TATTCATCGC CTTGAAAAAG CTGGTTTGGC TAAACGAGTT TTATTCTTAG CAGATAGAAA	1560
CATCTTAGTA GACCAAACGA TGGCTGAAGA CTTTAGGCCA TTCGAAAAGG TAATGACGAA	1620
AATTACACCA AAACTTTGA CTGCTCTGA AAAATTAAAT TCTTTTGAAA TTTATCTAGG	1680
GCTTTATCAG CAACTAACTG GTGAAGATGG AACTGAAACA CATTATCAAA AATTTGACAA	1740
AGACTTCTTT GATTTAATCG TAATTGATGA AGCGCACCGT GGTTCAGCTA AGGAAAACAG	1800
TAACTGGCGT AAGGTAATTG ATTATTTGAG TTCTGCGACA CAGATTGGGA TGACCGCTAC	1860
TCTTAAAGAA ACCAAGAATG CTTCCAATAC GGAATACTTT GGTGAGCCAA TCTATACTTA	1920
TAGTTTAAAA CAGGGAATCG AGGATGGTTT TTTGGCTCCA TATCGTGTTA TGAGGGTTAA	1980
TTTAGATGTG GATGTGGATG GTTATCGTCC AGAACTGGA AAAGTTGATG CTAACGGACA	2040
ATTAATAGAA GATAGTACT ACGGCAGGAA AGATTTTGAT AAAACCATTG TCATTGATGA	2100
TAGAACGCAA AGAGTTGCCA AGTTTGTTTC TGATTATATG AAGCAAAACA ATGCACGATT	2160
TGATAAAACA ATTGTTTTTT GTGTTGATAT TGACCATGCC GAGCGAATGC GTGCTGCACT	2220
TGTAAAGAG AATCTAGACT TAGTCCAAGA AGACTATCGT TATGTCATGC AAGTAACTGG	2280
TGACAACGCT GAAGGAAAAG CTCAACTGGA TAACTTTATG GATGTCAATT CTAATTTTCC	2340
CGCTATGTGA ACAACGTCTA AATTATTAAC GACAGGAGTT AATGCTAAAA CATGTCGTTT	2400
GATTGTTTTA GACTCTAATA TCCAATCCAT GACTGAATTT AAACAAATTA TTGGTCGTGG	2460
CACACGTCTT TATCCTCAAA AGGGGAAAGA ATTTTTCACG ATTATTGATT TTCGAAATGT	2520
TACCAATTTG TTTGCTGACC CTGATTTTGA TGGTGATCCA GTGAAGGTGC TAGAAACAGG	2580
TGCGAAAACA GTCAGTGGTT CTACGCCCGG TTTCGTAGAT GAGGAAGGTG ACCCAGTAGA	2640
AAAAATATATC GTTACAGACA AGCAGGTTAC CATTCTTAAT TCTACTGTTT AAGTATTGGA	2700
TGAAAACGGG AAACGTGATTA CCGAAAGCCT GACCGACTAC ACTCGAAAGA ATATCTTAGG	2760

1011

TAGCTACGCC ACTTTGAACG ATTTTATCAC AGTTTGGCAT ACGGCAGATA AGAAGAAGCT	2820
TATCTTAGAC GAACTTTATA AAAAAGGAGT TTATCTAGAT GCTATTGAG AGTCGGAGGG	2880
AATATCAGAA CAAGAAATCG ATGATTTTGA TTTACTCCTA AAACCTGCCT ATGGTCAAAA	2940
AGAATTAACC AAAACGGAAC GTATCAATAA ACTCAAACAA AGCGGATATT TATATAAATA	3000
TAGTGAGGAA GCGCGTGCTG TTTTGAAAT TTTACTGAAC AAATACATGG ATAAAGGTAT	3060
<hr/>	
TGGAGAACTC GAAAGCATTG AAACATTAAA ACTTCCAGAA TTTCAGATAT ATGGTGGAAC	3120
CTTCAAAATC ATCAATACTT ATTTTGGAGA TAAAAACGA TATTTACAAG CAATTAAAGA	3180
ATTGGAGCAA GAGCTATTTA CAGTAGCTTA ATGAAAGGAA AGTATGTCAA TTACATCATT	3240
TGTAAAAAGA ATTCAAGATA TCACTCGAAA CGATGCTGGT GTTAATGGTG ATGCTCAACG	3300
TATTGAGCAA ATGCTTGGT TATTATTCTT AAAAATTTAT GATAGCCGTG AAATGGTTTG	3360
GGAATTAGAA GAAGACGAGT ATGAGTCAAT TATCCCAGAG GAATTAATAAT GCGGAAATTG	3420
GGCTCATGCT CAAAATGGGG AACGGGTATT GACAGGCGAT GAATTACTTG ATTTTGTCAA	3480
TAACAAGTTA TTCAAAGAGT TGAAAGAGCT TGAAATAACT TCAATATGC CTATTCGAAA	3540
AACGATTGTT AAATCAGCTT TTGAAGATGC GAACAACTAT ATGAAAAATG GCGTCTTGTT	3600
ACGCCAAGTC ATCAATGTTA TTGATGAAGT TGATTTCAT AGCCCTGAAG ATCGTCATTC	3660
GTTTAATGAT ATTTACGAAA AAATTCTTAA AGATATTCAA AATGCTGGGA ACTCAGGAGA	3720
ATTTTATACG CCACGTGCAG CGACTGATTT TATTGCCGAA GTTCTTGACC CAAAACCTGG	3780
AGAATCAATG GCAGACCTTG CTTGCGGAAC AGGAGGCTTC TTGACTTCGA CTCTGAACCG	3840
TTTAAGTAGT CAACGTAAAA CTAGTGAAGA TACCAAAAAA TATAATACAG CTGTTTTTGG	3900
TATTGAAAAG AAAGCATTTT CTCATCTTTT AGCAGTTACA AATCTGTTT TTCACGAAAT	3960
TGATGACCCT AAAATTGTTT ATGGAAATAC TTTGGAGAAA AATGTTGCTG AATATACGGA	4020
TGATGAAAAA TTTGACATTA TTATGATGAA TCCACCTTTT GGAGGGTCAG AATTAGAAAC	4080
AATAAAAAAT AACTTTCCAG CAGAATTACG GAGTTCTGAA ACAGCTGATT TATTTATGGC	4140
TGTCATTATG TATCGTTTGA AAGAAAATGG TCGTGTGGA GTTATTTTAC CTGATGGTTT	4200
TCTATTTGGT GAAGGTGTAA AAACCTGCTT GAAACAAAAA CTGGTAGATG AGTTCAACTT	4260
GCATACGATT ATTAGGTTGC CTCATAGTGT CTTTGCACCG TATACAGGAA TCCATACGAA	4320
CATTCTTTTC TTTGATAAAA CAAAGAAAAC AGAAGAACT TGGTTTTATC GTTTAGATAT	4380
GCCAGATGGT TATAAAAAAT TCTCGAAAAC TAAGCCGATG AAGTCAGAAC ACTTCAATCC	4440
TGTTCTGAC TGGTGGGAAA ATCGTGAAGA GATTCTGGAA GGTAAATTCT ACAAATCTAA	4500

1012

ATCATTTACA CCTAGTGAAT TGGCTGAGTT GAATTATAAT TTAGACCAGT GTGACTTTCC	4560
AAAAGAGGAA GAGGAAATCT TAAATCCCTT TGAGTTGATT CAGAATTATC AAGCGGAAAG	4620
AGCAACTTTA AATCATAAGA TTGATAATGT ATTAGCTGAT ATTTTGCAGT TGTGGAGGA	4680
CAAATAATGA CACCAGAACA ACTTAAAGCA AGTATTCTCC AAAGAGCGAT GGAAGGGAAA	4740
TTAGTGCCGC AAAATCCCAA TGACGAACCT GCAAGTGAAT TATTAAAGAG AATTAAAGCT	4800
GAAAAAGAAA AACTTATCAG TGAAGGAAAA ATCAAACGAG ATAAAAAGGA AACTGAGATA	4860
TTTCGTGGTG ATGATGGGAA ACATTATGGG AAGTTTGCTG ATGGAAGCAC TCAAGAAATT	4920
GATGTTCCCTT ATGATATTCC TGATACTTGG GAGTGGGTGA GGTTTTCTAC ATTGGTTGAA	4980
ATTGTCAGAG GTGGCTCTCC ACGACCAATC AAAGATTATC TTAATTCTGA AGTAGATGGA	5040
ATAAATTGGA TAAAAATAGG TGATACTGAA AAGGGTGAAA AGTATATAAA TAATGTTAAA	5100
GAAAAAATCA AAAAATCAGG GCTTAACAAA ACTAGATTTG TAAAAAAGG TACATTTTGT	5160
TTAACTAATT CTATGAGTTT TGGTAGACCT TATATTTTGA ATGTTGATGG TGCAATACAC	5220
GATGGATGGT TGGCTATTTT GAACTATGAA AACTCATTA AATAAGATTA CCTATTCTAT	5280
ATTCTTTTAT CAAATGTAGT TTATTCTCAA TTTCTATCTC TAATTAGTGG AGCTGTTGTG	5340
AAAAACTTGA ATAGTGATAA AGTTGCTTCT ATTCTTATCC CTCTCCCCC ACTATCCGAA	5400
CAACAACGAA TAGTAGAAGC AATCGAATCA GCTTTAGAAA AAGTAGATGA ATATGCTGAA	5460
AGTTATAATA GACTAGAACA GCTAGATAAA GAATTTCCAG ATAAACTAAA AAAATCTATT	5520
CTTCAATATG CTATGCAAGG AAAATTAGTT GAACAAGACC CAAATGATGA ATCAGTCGAA	5580
GTTTACTTGT AAAAAATACG AGCAGAAAAA CAAAACTCT TTGAAGAAGG CAAGATTAAA	5640
AAGAAAGATT TGGACATTTT TATTGTTTCC CAAGGAGATG ATAACTCTTA TTATGGGAAT	5700
ATACCTATGA ATTGGGTTGT TATAAAAATA AAAGATATTT TTTCAATAAA TACAGGTCTT	5760
TCTTACAAGA AGGGCGATTT AAGCATTAAT AATAAAGGTG TTAGAATTAT ACGTGGTGGT	5820
AATATTAAGC CTTTGAATT TTCTCTGTTG GATAATGATT ACTACATTGA TACACAATTC	5880
ATCTCCTCTG AGCAAGTTTA TTTAAACAT AATCAGCTAA TAACACCTGT ATCAACCTCT	5940
TTAGAACATA TTGGAAGTT TGCAAGAATC GATAAAGACT ATGATGGTGT TGTGGCTGGT	6000
GGATTTATTT TCCAATTAAC ACCATTGCGAA AGTTCAGAGA TTATTTCAAA ATTTCTATTA	6060
TTTAACTTGT CCTCTCCGTT ATTTTATAAA CAATTGAAAG CAATAACTAA ACTATCAGGT	6120
CAAGCTTTAT ATAATATTCC TAAACTACA CTGAGCGAGC TATTAATTCC GTTAGCTCCT	6180
TTTGAGGAAC AGGAACCTAT TACTCAAAAA GTTGAGAAAC TTTTGA AAA AGTAAATCAA	6240
CTTTGAAAAT GATTCTTTT ATCTCTTCAT GATTAGAAAT AGGGATTAAT AATTCGAGAA	6300

1013

TACTGGTACT ATTTAATGTT TTCCCTTTGA TAGCATCTTT TGAATCACCT AAAGTAGAGA	6360
TAAGTGGCAA AAATATCATT AAGTAATCTC TGATAATATT TTCTTTATTA GCATAGGGGA	6420
ATATCGATAT AATGGCTTCA TTATGAGTGG CAGGAATATC CAATATGGCA ACTTTTCCAA	6480
TAGATAATTT AAAACTCATT AATAAAGTTC CTTTAGGTGA AATGTCTATT TTCTTTGATT	6540
TTAATGCTAA TTTAGAAATA GATTCTCTCG CATTAGTTAC ATAACCAGAT ATAGGCATAT	6600
<hr/>	
CTGATATAGA TACCCAAGGT ATTTCAAGTC CCCAAAAAGT AGCTTCACTG CGTGGAGGAG	6660
TTTTTCCTAT TCTGAAGTTA ACTAGGCTAG CAAATTTAAT ATATCTCCAT GCTTCTGGGA	6720
TTTCATATAT AGGATAAGAG GTTGTTCGT CTTGTTCCTC ATAATAAGAG CCATAATCAC	6780
AAAAATAGCA GGTAGTCAGT TTGACCACCT GTTATTTTTT ACCAATTAAC AATTTTATCT	6840
ACAATATTTT GTTGTTCAGT AGCTGTTTC CTTAGATAAA TTCGAGTAGT TTCTATACTT	6900
TCGTGTCCCA TCAAATCTGC AAGCAAGGCA ATATCATTAT ACTTCGCTAA AAAATTCTTA	6960
GCAAATAAAT GCCTAAAAGA ATGAGGGTAA ATTACGTTAG GATTCATTTT GTATTTATCA	7020
GCATAATTTT TTAAGTGTG AGCAACTCCT CTTGCTGTAA TTGGTTCGTT AAATTTATTC	7080
AAAAATAAAT AACCCTTCG GCGATTTTCT GATTCTAACC AACTAAGACA ACTATTTCTT	7140
AATTTTTTAG GAATGTACAG TCTACGAATT TTACCACCTT TTGAGTAAAT GTCAAAATAA	7200
CCGATTTCTA CATGCTCTAC TTTTAGTTTA ATAAGTTCAC TTACACGAGC CCCAGTTGCA	7260
CCTAAAAACC AAACGACAAA ATGCCATTTT AAAATACCAT CTTTTTTCAA ACTACGTTTA	7320
AGAAAAAGGT AATCAGCATG GCTAATGACA TCTTCTAAAA ACGGTTTTTG CTGTACTTTG	7380
ACAAATTTTA ATTTCAAATC ATCATGACCA ATAAAAGCCA GATATTTATT TACTCCTTGT	7440
AGTCGCAAAT TGACAGTTTT AGGTTTAAAA TTGTCTAATA AATATCCTTT GTATTCAAAT	7500
AAATCTTCCA TTTTGAGTTC GTAATCTCTC AAGAAAAATC GAACACCATA AAGGTACGAA	7560
CGCACAGTAT TTTCAGCTAA ACCAGCTTTC TTCAAATGTA ATTCAAATC TTCAACGTA	7620
AACTCCTAT CTTATGTTTG ATAGAAATTC CACCGCACGT AAACTATTA TACTAAATTA	7680
GTGCGTCAAT ATGGGCGAAA AATTGTTTGA TTTTATCAAC GATTCTGGAT TGTTACAGAA	7740
GGGGTGGGAG GGGGATTAAA TATTCTTTTA TAGTTTTCTG TAATAATTCT TTTGTTTTG	7800
TACTACCCGA CGTTTTTCT TCAATAACTG ACTGAACAAT AGGAGAGGAA AGAAAATTAT	7860
AGATGAAATG GCAATTAATA ACCCCCGATA AGACTCTTAT AACTGTAACA TGGCTATCTG	7920
CAACAGCCCA GCCATAAGGA TTTTATTTT CATGGTAAAT AGCTAATCGT CCTAACGTAC	7980
CTAGACCTGT TGAATTCAC ATTAAATCAC CATCTCTTAG TAATCTTCT TTCTGGTAAC	8040

1014

TATGAACTGT TTCGGGATCA ATAAATCTTG CTAAGTCAAT AGAAAAGCCA GACCATTGAT	8100
TACATTTCTG AGCAATCACA GGGTATATAG GAATATTTGA ATATTTTGGA GACTTCCCTC	8160
TTTGAATGTA GGAGGTTATA TCGTTTAACC TCACCCATTC CCAACTTTCT GGTATTTTAC	8220
AAGGTACTTC CTCATAATAA GAGTTATCAT CTCCTTGGGA AACAATAGAA ATGTCCAAAT	8280
CTTTCTTTTT AATCTTGCCT TCTTCAAAGA GTTTTTGTTT TTCTGCTCGT ATTTTTTCAA	8340
GTAAACTTTC GACTGATTCA TCATTTGGGT CTGTGTTCAAC TAATTTTCCT TGCATAGCAT	8400
ATTGAAGAAT AGATTTTTTT AGTTTATCTG GAAATCTTT ATCTAGCTGT TCTAGTCTAT	8460
TATAACTTTC AGCATATTCA TCTACTTTT CTAAGCTGA TTCGATTGCT TCTACTATTTC	8520
GTGTGTGTTT GGATAGTGGG GGGAGAGCAA TTAATAATAG ATTAAATTA TAATCATTGA	8580
TTGCAGGATA ACTTGTTCCTA GTAGATTTAT TATTAACACG ATTGATAAAA TTATCTGATA	8640
ATAAATAATA TTTCAAATAT GTTTCGTAA GTAAAGTATC CAAAACAATA AATGCTGTAC	8700
TAGCTATCAA ATACTCTTTA AGTTCTCTAA CTACAGCAAT ATTTTTTAGA TATGGTCTAA	8760
CTGTGAAAA TAAGACACTA TTCTGCGAAA CTAATTTTCT AGCACGGGAA GGCGCTTGTT	8820
CAGGTGAAAG ATATTGTAGA TTTTGTAGT TGATTATGTT CTTTTTCTA TCAATACTAG	8880
ACGTATCTAT ATACCTAAAG GATTTCTCTG GCTTATTTTG CCCAAAATTC CAATAAATTG	8940
ATTTTATCCT CACCCACTCC CAAGTATCAG GAATATCATA AGGAACATCA ATTTCTTGAG	9000
TGCTTCCATC AGCAAACTTC CCATAATGTT TCTTATGTGC TTCAAGTATA TAAAAAGGCG	9060
TAAAAATACG CCTATAGATA ATGGGGTTGA AATAGGTTTA TTGTTGATGA GATTGTAGAT	9120
AATCAATTT TTTACTTCCA ATCGAATATT CAAATCCTCC ACCTTTTCTG CCTGTAATTG	9180
TTCATCATAA AATTCAATAT CTTCAGGATT TTCCCTTGG CAACCTCGGC AGAAATATTC	9240
TTCCGCTCGA TCAGGATTCA AAAATCGACA AGCACAACA AACAGTCGC CATCATCATT	9300
TATTGAGATA ATATAGTAGA TTGAAATAAG ATGTAAACAA ATCGATTAGG AAAGTTAAAT	9360
TAGTTTCTAG AAATTTTtag CAGATGTAGT GTACTATTCT AGTCTCAATT TACTATGGCT	9420
TCAAATATAT CTTTCGAAAA AATATTTTACA GATGTGTAAT TTTGAAGCTT GCAAAAGTTA	9480
GTAAACTTGT AGATTTTCAT TTGAAGTAAC TTGTTTCTT GCCCGATATT GTTTTTGAAA	9540
TTGAATTTT CCATAGTGAC TCCTTAATTT TCTTCTACAC GTCTGATGAT AAATCTAATT	9600
CGCAAAGAG TCAAGAGGAT TTTTCGAAAA ATAAATAGCG ACCGAAATCG CTATTTTAAG	9660
GGTTATAGGT ATTTGATGGC TTAGACTGCT GTGTGACTGT TTACCCACAG GCAATCTTTC	9720
TTCTATATTA GTATTAGTAA AGGTCTAAAT AATTATCAAT TTCCCATTTG GAAACGAAGG	9780
TTGCATAACT TGCCCATTCG ATTCGTTTGG CTTCAAGGAA GCTAGTATAG ATGTGATCTC	9840

1015

CGAGAGCAGC TTAAACCACT TCATCTTCTG TCAAAGCTTT CAAAGCGTTG TGAAGAGTTG	9900
ATGGAAGGTC TGTAATACCA GCTTCCTTGC GCTCTTCTGC TGTCATGATG TAGATATTTT	9960
CTTCGATAGG AGCTGGTGCT TCGATTTTAT TTTCAATACC ATACAAACCA ACTTCCAAAA	10020
GAACAGCCAT AGCAACGTAA GGGTTCGCCA TTGGATCCAC TGAACGCAAC TCAAGACGAG	10080
TTCCCATACC ACGTGAAGCA GGTACGCGCA CAAGTGGCGA ACGGTTACGA CCAGCCCAAG	10140
CAATGTAAAC AGGCGCTTCA TAACCTGGAA CCAAACGTTT GTATGAGTTA ACTGTTGGGT	10200
TCATGATGGC AGTATAGTTG TAAGCATGCT TGATCAAACC GCCTAGGAAA TGGTAAGCTG	10260
TTTCTGACAA CTGCATTCTT TTTGGATCAT TTGGATCAAA GAAGGCGTTA TTTCTTCTG	10320
CATCAAACAA GGACATATTA CAGTGCATAC CTGATCCAGC AATACCAAAT TTTGGCTTCG	10380
CCATAAATGT TCGGTAAAGT CCGTGTCTGC GAGCAATGGT TTAAACAACA AGCTTAAAGA	10440
TTTGAATCTT ATCACAAGCA CGGAGAACTT CATCGTACTT AAAGTCAATC TCATGCTGTC	10500
CAACCGCAAC CTCGTGGTGA CTCGCTTCTA CTTCAAATCC CATTTTGGTC AAGACATTCA	10560
CAATCTCAGC ACGTGTGTTG TCCGCAAGGT CAGTAGGTGC CAAGTCAAAG TAGCCACCCT	10620
TGTCATTAC CTTCAAGTGT GGGTCCCAT TTTCATCCAA CTTAAATAGG AAGAATTCTG	10680
GCTCTGGACC AAGGTGAAG GATTGAATC CAACTTCTTC CATGTGACGA AGAGCTCGTT	10740
TCAAATTACC ACGAGGTCA CCCGCAAATG GTTCACCTTC TGTGTATAG ACATCACAGA	10800
TCAGACCTGC AACACTTCCA TTTCATCTC CCAAGGGAA GACTGTCCAT GTATCCAAGT	10860
CCGGGTACAA GTACATATCC GACTCATTGA TACGTACAAA ACCTTCAATA GAAGATCCAT	10920
CAAACATAAC CTTGTTCGAC AAGACCTTAT CTAAGTCTC ATCTGTAGCA GGAATTCGA	10980
CGTTTTTCAT GGTTCCTCAA ATATCTGAGA ACATAAGACG AATAAAGGTA ACATTTTTTT	11040
CCTTGACTTC ACGACGAATA TCTGCAGCTG TGATTGGCAT AAGTTTTCTC CTTAATCTAT	11100
GACTACTTGC GGTTCCTTAA CCGCGACCAA AAGGTGACTG TACTGAAGCA AAACGCCCCT	11160
GTTGGAGGAG TTCATTGTGA AGTGCACGAC GACTTTCAGT CTGACTAACC GCTTCTTGG	11220
ATTTTCGCTC ACGTTCAGCA TATTTTTTCT TAATGGCAGC GATATTATAA CCTTCAGAGA	11280
TATAATCTTT GATTTCAAGC AGACGATCCA TGTCATTCAA GGAATACATG CGACGATTC	11340
CTTCGTTTCG ATCGGGCTTG ATCAACTCTT GATCTTCATA ATAACGAATC TGACGCGCCG	11400
ATAGATCGGT CAACTTCATA ACACTGCCGA TAGGAAAAAC AGCCATATTT CGGCGAAATT	11460
CTTTTTCTT CATTTACAAT TTCCTTCTTT CTGTCTATTA TAGTCTAAAA AAAGACAAAC	11520
GTCAATTGAT AATGTTATAA AATGTAACAT TATTTTTCTT TTTTCTCTAA AAAGAGACGA	11580

1016

ATACGATCAA TATCGTAATT TACGATAATT GCGACAAAAA CTCCCATAAA CGTTTCTAAT	11640
ACACGCACAA ACACGTACAA AATTGTCTCA CCACTTGGAA TTGATAGGGT AATGATTAAC	11700
ATAGCTGCTA CACCACCAAT AACCCCTGCT TTGTTATTCA TGGCTACATT TGTGATAATG	11760
GTTAACATGG TGCAGATTGG AACAACTACC AAGGTCACCC AAAAGGCTTC GTGGAAAAAG	11820
GTATTTAATA AGAAGAAGAC CAAGGCATAG AGTCCACCGA TACTATTTCC TAGAATACGC	11880
GAAGTCCCAA AATGAACACT CTCATCAAAA CTCTCCCTCA GGCTAAAAAC GGCTGTCAAA	11940
GCACCAATTT GAAGACCTTT CCAGCCAAAA AAGCCAAAA TCAAGAGAAC TAGAAAAACA	12000
GCAATACCTG TTTTAAAGT TCGCATACCA AGTTTGAAGT GGGATTTATC GAATTTATAT	12060
TTTTTAAAT AACTCATAAT CTCAACTTTC TATTTCCATT TTATCATAAA TCGGTGATTT	12120
TTATGAGTAA TAGTTGAGAG GAAGCGTTTT TATTTTAAGC AAAAGAAAAG AGGAACTTTC	12180
ATCCCTCTCT TCTTTGATTT ATTTATAAAA TCTTATTTTT CTGTCAAGGC TGCAAGTCTT	12240
GGAAGAACCT TACCTTCAAG AAGTTCCATT GATGTCCAC CACCCGTAAT AATCCATGAG	12300
AACTTGCTCG CACGGCCAAG GTTAATCGCT GCGGCAGCTG AGTCACCACC ACCGATGATT	12360
GATTTAACTC CTGGTTGTTT CACGATAGCG TCCATCACAC CGATTGTACC AGCTTGGAAA	12420
TCTGGGTTTT CAAATACACC CATAGGTCCG TTCCATACGA CTGTTTTGGC ACCAGTCAAA	12480
GCTTCGTCAA ATTTGGCGAT AGATTTTGGA CCGATGTCAA GACCAAGGAA GCCTTCAGAA	12540
ACTGCTTCAC CTTCAGTGTC ACGCACTTCA GTGTAACCAG CAAATGCGTT AGCTTCTTTT	12600
GAGTCAACTG GCAAGATCAA TTTACCATTT GCTTTTTCAA GAAGAGCTTT CGCAACATCC	12660
AATTTGTCTT CTCTACAAG TGAGTTACCG ATTTGATAC CTTGTGCTTT GTAGAATGTG	12720
TAAGTCATCC CACCACCGAT AAGGACGTTA TCAGCTTTTT CAAGCAAGTT TTCGATAACA	12780
CCGATCTTGT CTGAACTTT TGAACCACCA AGGATAGCCA CGAATGGACG TTCTGGAGTT	12840
TCAACTGCTT CTTGGATGTA GGCAATTTCTG TTTTCAAGAA GGAAACCAGC AACTGCTTTT	12900
TCAACGTTTG CTGAGATACC AACGTTAGAT GCGTGTGCAC GGTGAGCTGT ACCGAATGCA	12960
TCGTTTACGA AGATACCATC TCCAAGTGAT GCCCAGTATT TACCAAGTTC AGGATCGTTT	13020
TTAGATTCTT TCTTGCCGTC AACATCTTCG TAACGAGTGT TTTCAACCAA GAGAACTTGT	13080
CCATCTTCAA GAGCGTTGAT TGCCGCTTCT AATTCAGCAC CACGAGTGAC ACCTGGGAAA	13140
ACAACATCTT GACCAAGTTT TGCTGCCAAG TCAGCTGCTA CAGGAGCAAG TGATTTACCA	13200
GCTTTATCAG CTCTTCTTT CACACGTCCA AGGTGAGAGA AAAGAATTGC ACGTCCACCT	13260
TGTTGATGA TGTACTTAAT AGTTGGAAGA GCTGCTGTGA TACGGTTATC GTTAGTGATT	13320
ACGCCATCTT TCAATGGTAC GTTGAAGTCA ACACGAACGA GGACTTTTTT ACCTTCAAG	13380



1017

TCAACGTCTT TAACAGTAAG TTTTGCCATG TTACAAAAAC TCCGG

13425

(2) INFORMATION FOR SEQ ID NO: 152:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 905 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

GATTTATCCT ACCGGnGAAT TTCCGGAGGG GTTCTAGCAG CAATCTTAGG AATCTATGAA	60
CGAATGATTG GCTTTCTGGC CCATCCCTTT AAAGACTTTA AAGAAAATGT TTTGTACTTT	120
ATTCAGTTG CCATCGGTAT GCTTCTGGGA ATCGGCTTAT TTTCTACCC GATTGAATAC	180
CTGCTTGAAG ATTATCAGGT TTTTGTATTA TGGAGCTTTG CGGGAGCTAT TATCGGTACA	240
GTTCCTAGCC TCCTCAAAGA ATCAACTCGA GAATCTGACC GAGACAAGAT TGATTTAGCT	300
TGGTTATGGA CAACCTTTAT CATTCTGGA TTAGGACTCT ATGCCTTAAA TTTTGTGTT	360
GGAACCTTAA GCGCCAGCTT TCTTAACTTC GTCCTAGCAG GCGCACTATT GGCCCTTGGC	420
GTCTTGGTTC CTGGCCTCAG CCCATCAAAT TTACTTTTGA TTTTGGGACT CTATGCTCCT	480
ATGTTGACTG GTTTTAAAC TTTTGATTTT TGGGAACTT TCTTTCCGAT TGGAATTGGT	540
GCAGGTGCAA CTCTCATCGT TTTTCAAAA TTGATAGATT ATGCCTTAAA CAACTACCAC	600
TCACGCGTCT ATCATTTCAT CATCGGTATC GTCCTATCAA GTACCCTTTT GATCTTAATT	660
CCAAATGCAG GAAACGCTGA AAGTATCCAA TACACAGGAC TTTCACTTGT CGGTATGTC	720
ATCATCGCCT TCTTCTTTGC GCTGGGAATC TGGCTTGGTA TTTGGATGAG TCAATTGGAG	780
GATAAATATA AATAATGGCA AAAAAAGTTA AAATCAAAAA AACATTGGTG GAACAAATCC	840
TATCTAAAGC AGCTATCCCT CATCAGGGGA TTCAAATCAA TGCCCTAGAA GGAGAGCTTC	900
CTCAA	905

(2) INFORMATION FOR SEQ ID NO: 153:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4278 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

1018

CTTGAATTAA ATAAAAAACG TCATGCGACT AAGCATTTTA CTGATAAGCT TGTGATCCC	60
AAAGATGTGC GTACGGCTAT CGAAATTGCA ACCTTAGCGC CAAGCGCCCA CAACAGCCAG	120
CCTTGGAAAT TTGTGGTGGT ACGTGAGAAA AATGCTGAAC TGGCAAAGTT AGCTTATGGT	180
TCCAATTTTG AACAGGTATC ATCAGCGCCT GTAACCATTG CCTTGTTTAC AGATACGGAC	240
TTAGCCAAAC GTGCTCGTAA GATTGCCCGT GTTGGTGGTG CTAATAACTT TTCTGAAGAG	300
CAACTTCAAT ATTTTATGAA AAATCTGCCA GCTGAGTTTG CCCGTTACAG TGAGCAACAA	360
GTCAGCGACT ACCTAGCTCT CAATGCAGGT TTGGTTGCCA TGAAGTTGGT TCTTGCATTG	420
ACAGACCAAG GAATTGGTTC TAACATTATT CTGGTTTTG ACAAATCAAA AGTTAATGAA	480
GTTTTGGAAG TCGAAGACCG TTTCCGCCCA GAACTCTTGA TCACAGTGGG TTATACAGAC	540
GAAAAATTGG AACCAAGCTA CCGCTTGCCA GTAGATGAAA TCATCGAGAA AAGATAGAAA	600
GAAGAAAAAA TGACAGCAAT TGATTTTACA GCAGAAGTAG AAAAACGCAA AGAAGACCTC	660
TTGGCTGACT TGTTTAGCCT TTTGGAAATC AATTCAGAAC GTGATGACAG CAAGGCTGAT	720
GCCCAGCATC CATTTGGGCC TGGTCCAGTA AAAGCCTTGG AGAAATTCCT TGAAATCGCA	780
GACCGCGATG GCTACCCAAC TAAGAATGTT GATAACTATG CAGGACATTT TGAGTTTGGT	840
GATGGAGAAG AAGTTCTCGG AATCTTTGCC CATATGGATG TGGTGCCTGC TGGTAGCGGT	900
TGGGACACAG ACCCTTACAC ACCAACTATC AAAGATGGTC GCCTTTATGC GCCCGGGGCT	960
TCGGACGATA AGGGTCCTAC AACAGCTTGT TACTATGGTT TGAAAATCAT CAAAGAATTG	1020
GGTCTTCCAA CTTCTAAGAA AGTTCGCTTC ATCGTTGGAA CAGACGAAGA ATCAGGCTGG	1080
GCAGACATGG ACTACTACTT TGAGCACGTA GGACTTGCCA AACCAGATTT CGGTTTCTCA	1140
CCAGATGCTG AATTTCGAAT CATCAATGGT GAAAAAGGAA ATATCACGGA ATACCTCCAC	1200
TTTGCAGGAG AAAATACAGG TGTTGCCCGT CTTACAGCT TTACAGGTGG TTTACGTGAA	1260
AATATGGTAC CAGAATCAGC AACAGCAGTC GTTTCAGGTG ACTTGGCTGA CTTGCAAGCT	1320
AAACTAGATG CCTTTGTTGC AGAACACAAA CTTAGAGGAG AACTCCAAGA AGAAGCTGGC	1380
AAATACAAGG TGACGATCAT TGGTAAATCA GCCCAGGTG CTATGCCTGC TTCAGGTGTC	1440
AATGGCGCAA CTTACCTTGC CCTCTTCCTC AGCCAGTTTG GCTTTGCTGG TCCAGCCAAA	1500
GACTACCTTG ACATCGCAGG TAAAATTCTC TTGAACGATC ATGAGGGTGA AAATCTTAAG	1560
ATTGCTCATG TGGATGAAAA GATGGGTGCT CTTTCTATGA ATGCCGGCGT CTCCACTTC	1620
GATGAAACAA GTGCTGATAA TACCATTGCC CTCACATCC GCTATCCAAA AGGAACAAGT	1680
CCAGAACAAA TCAAGTCAAT CCTTGAAAAC TTGCCAGTTG TTTCTGTTAG CCTGTCTGAA	1740
CACGGTCACA CGCCTCACTA TGTGCCAATG GAAGATCCAC TTGTGCAAAC CTTGTTGAAT	1800

1019

ATCTATGAAA AACAACTGG CTTTAAAGGT CATGAACAAG TCATCGGTGG TGGAACTTT	1860
GGTCGCTTGC TAGAACCGGG AGTTGCCTAC GGTGCTATGT TCCCAGACTC GATTGATACC	1920
ATGCACCAAG CCAATGAATT TATCGCCTTG GATGATCTTT TCCGAGCAGC AGCAATTTAT	1980
GCCGAAGCTA TTTACGAATT GATCAAATAA AACGATAGAA GTCTGAGATC TTATGCTTGG	2040
ACTTCTTTTT GGAGGGAAAG TAGATGTCTC AAATCGAAAG AATCAAACAG GCTATCATGG	2100
CGGATTCGCA GAATGCCAGC TATACAGAGC GTGGCATTGA GCCTCTCTTT GCAGCGCCAA	2160
AAACTGCTCG CATCAATATC ATCGGTCAGG CTCGGGACT TAAACTCAA GAAGCAGGCC	2220
TTTACTGGAA AGATAAAGT GGTGACCGCT TCGGGGACTG GCTAGGTGTG GATGAAGATA	2280
CCTTTTACAA TTCAGGTTAT TTTGCTGTTT TGCCTATGGA TTTCTACTTT CCAGGACATG	2340
GCAAGTCGGG TGATCTTCCG CCTCGTACAG GTTTTGCAGA AAAATGGCAT CCGCAGGTCT	2400
TACAGGAATT GCCTGATATT CAGTTAACCC TCTTGATTGG GCAATATGCC CAAGCCTACT	2460
ATTTACAGGA GAAATCAGT GGGAAGGTAA CGGAGAGGT GAAACACTAT AAAGACTATC	2520
TGCCAGCCTA TTTTCCGCTA GTTCACCCAT CACCACGAAA TCAAATCTGG ATGGCCAAAA	2580
ATCCTTGGTT TGAGGCAGAA GTAGTGCCAG ATTTGAAAAA AAGAATTAAA ACCATTTTAT	2640
AGTCAATGAA AATCAAAGAG CAACTAGGA AGCTAGTCGT AGGCTGCTCA AAGTACAGCT	2700
TTGAAGTTGC AGATAAACT GACGAAGTCG GTAACATACG CACGGTAAGG CGACGCTGAC	2760
GTGGTTTGAA GAGATTTTCG AAGAGTATTA GAAGAAAAAG AATGAAAGAA ATAGCCTTTG	2820
ACGCATTTTA CCAGCTTTAC CAAAACGACC AGCTTTCTTT AGTGGATGTG AGAGAAGTGG	2880
ATGAGTTTGC AGCTCTTCAT TTAGAAGGTG CCCACAACCT ACCGCTTAGT CAATTGGCTG	2940
ATAGTTATGA TTAATTGGAC AAAGATCGCT TGCATTATAT TATTTGCAA TCTGGAATGA	3000
GATCGGCGCG TGCTTGCCAA TTCCTATTAG AACAAAGTTA TAATGTTATC AATGTCCAGG	3060
GTGGCATGTT AGCCTTTGAA GAACTTTAAA ATTTTGCAAT TCTCCTACTT GGTGTGGACT	3120
GGGTAGGAGA GTTTTATTTT TAGATAATTC TTATTTTAA GAAAATTGAA AACATTTAAT	3180
ATTTGCCTCG TGATGCTTTT TTCAGACTCC TAATCGTGGT ATACTAGGTC AGTATTTTAT	3240
AAATATGAAG GAGATTTTGA TGGCTAAAAA AGGTACCCTA ACAGGTTTGC TCCTGTTTGG	3300
AATATTTTTT GGTGCGGGGA ACTTGATTTT TCCGCTTCT CTAGGTGCTC TATCTGGAGA	3360
ACATTTTCTT CCTGCCATCG CAGGTTTTGT CTTTTCAGGC GTTGGTATCG CCGTCTTGAC	3420
CCTTATTATT GGAACGCTAA ATCTAAAGG ATATATCTAC GAGATTTCAA CGAAGATAGC	3480
GCCTTGGTTT GCGACTCTT ACCTCTCAGT TCTTTACTTG TCAATCGGTC CATTCTTTGC	3540

1020

TACCCACGT ACTGCTACAA CAGCTTACGA AGTAGGGATT AGCCCCCTTT TGTCGGATGC	3600
AAATAAAGGA CTTGGCTTGA TTGTATTTAC GGTCTGTAT TTTGCGGCAG CCTATTTGAT	3660
TTGCTTAAT CCATCAAAAA TCTTAGACCG CATTGGACGT ATTTTAACGC CAGTCTTTGC	3720
AATTTTGATT GTTATCTTGG TCGTTCTGGG AGCTATCAAA TATGGTGGAA CAAGTCCTCA	3780
AGCTGCTTCA CTGCTTATCA AGCTTCTGCC TTTGGTACAG GTTTCCTAGA AGGTTACAAT	3840
ACCTTGGACG CCCTTGCCTC AGTGGCCTTT AGCGTAATCG CAGTTCAAAC CTTGAAACAA	3900
CTTGGATTTT CAAGTAAGAA AGAATACATT TCAACTATTT GGGTTGTTGG TATCGTTGTT	3960
GCCCTTGCCT TCAGCGCTCT TTACATCGGT TTAGGTTTTT TGGAAATCA TTTCCCACTA	4020
CCAGCTGAAG CGATGAAGGG TGGAAACCA GGTGTTTACA TCTGTTCACA AGCCAACAA	4080
GAAATCTTTG GCTCAACAGC TCAACTCTTC CTTGCAGCTA TGGTTACCGT AACCTGCTTC	4140
ACAACGACTG TTGGTTTGAT TGTGTCAACA GCTGAGTTCT TTAATGAGCG CTTCCACAA	4200
ATCAGCTACA AGGTTTATGC GACAGCCTTT ACCTTGATTG GATTTGCTAT TGCCAATTTG	4260
GGTCTTGATG CGATTATC	4278

## (2) INFORMATION FOR SEQ ID NO: 154:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1953 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

ACCCGATCAA ATGACAAAAG CTAACCTTTG TGTCGTAGGT ATGGCCGTAA TGGGTCGTAA	60
CCTTGCCCTT AATATTGAAT CTCGTGGTTA CACAGTTGCT ATCTACAACC GTAGTAAAGA	120
AAAAACGGAA GATGTGATTG CTTGCCATCC TGAAAAGAAC TTGTACCAA GCTATGACGT	180
TGAAAGTTTT GTAAACTCAA TCGAAAAACC TCGTCGTATC ATGCTGATGG TTCAAGCTGG	240
ACCTGGTACA GATGCTACTA TCCAAGCCCT TCTTCCACAC CTTGACAAGG GTGATATCTT	300
GATTGACGGA GGAAATACTT TCTACAAAGA TACCATCCGT CGTAATGAAG AATTGGCAAA	360
CTCTGGTATC AACTTTATCG GTACTGGGGT TTCTGGTGGT GAAAAAGGTG CCCTTGAAGG	420
TCCTTCTATC ATGCCTGGTG GACAAAAAGA AGCCTACGAA TTGGTTGCGG ATGTTCTTGA	480
AGAAATCTCA GCTAAAGCAC CAGAAGATGG CAAACCATGT GTGACTTACA TCGGTCCTGA	540
TGGAGCTGGT CACTATGTGA AAATGGTTCA CAATGGTATT GAGTACGGTG ATATGCAATT	600
GATCGCAGAA AGCTATGACT TGATGCAACA CTTGCTAGGC CTTTCTGCAG AAGATATGGC	660

1021

TGAAATCTTT ACTGAGTGGG ACAAGGGTGA ATTAGACAGC TACTTGATTG AAATCACAGC	720
TGATATCTTG AGCCGTAAAG ACGATGAAGG CCAAGATGGA CCAATCGTAG ACTACATCCT	780
TGATGCTGCA GGTAACAAGG GAACTGGTAA ATGGACTAGC CAATCATCTC TTGACCTTGG	840
TGTACCATTG TCACTGATTA CTGAGTCAGT GTTTGCACGC TACATTTCOA CTTACAAAGA	900
AGAACCTGTA CATGCTAGCA AGGTGCTTCC AAAACCAGCT GCCTTCAACT TTGAAGGAGA	960
CAAGGCTGAA TTGATTGAAA AGATCCGTCA AGCCCTTTAC TTCTCAAAAA TCATTTCATA	1020
CGCACAAGGA TTTGCTCAAT TGCCTGTAGC CTCTAAAGAA AACAACGGGA ACTTGCCATT	1080
TGCAGATATC GCATCTATCT GGCCTGATGG CTGTATCATC CGTTCTCGTT TCTTGCAAAA	1140
GATTACAGAT GCTTACAACC GCGATGCAGA TCTTGCCAAC CTTCTTTTGG ACGAGTACTT	1200
CTTGGATGTT ACTGCTAAGT ACCAACAAGC AGTACGTGAT ATCGTAGCTC TTGCGGTTCA	1260
AGCAGGTGTG CCAGTGCCAA CTTTCTCAGC AGCTATTACT TACTTTGATA GCTACCGTTC	1320
AGCTGACCTT CCAGCTAACT TGATCCAAGC ACAACGTGAC TACTTTGGTG CTCACACTTA	1380
CCAACGTAAG GACAAAGAAG GAACCTTCCA CTACTCTTGG TATGACGAAA AATAAGTAGG	1440
TCAGCCATGG GGAAACGGAT TTTATTACTT GAGAAAGAAC GAAATCTAGC TCATTTTSTA	1500
AGTTTGAAC TCCAGAAAGA GCAGTATCGG GTTGATCTGG TAGAGGAGGG GCAAAAAGCC	1560
CTCTCCATGG CTCTTCAGAC AGACTATGAT TTGATGTTAT TGAACGTTAA TCTGGGAGAT	1620
ATGATGGCTC AGGATTTTGC AGAAAAATTG AGCCGAACCT AACCTGCCTC AGTCATCATG	1680
ATTTTAGATC ATTGGGAAGA CTGCAAGAA GAGCTGGAAG TTGTTACAGG TTTTGCAATT	1740
TCATACATCT ATAAGCCAGT CTTATCGAA AATCTGGTAG CGCGTATTC GGCGATCTTC	1800
CGAGGTCGGG ACTTCATTGA TCAACACTGC AGTCTGATGA AACTTCCAAG GACCTACCGC	1860
AATCTTAGGA TAGATGTTGA ACATCACACG GTTTATCGTG GTGAAGAGAT GATTGCTCTG	1920
ACACGCCGTG AGTATGACCT TTTGGCGACA CGG	1953

(2) INFORMATION FOR SEQ ID NO: 155:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6474 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

CCGGCAGTAC ACGAGCTTGG GGAACAGCCA CTGGAACGAT GAGGTGTGAG CTCAAAATAT 60

1022

CCTCCAGTTA TGTTTTTCCT AATAGTATAC CGGAAGAGTG AAAGGATTTT ATAATGGAGC	120
GGTTACAAAG AACCTACTTT CTATTAAACA GTATACTATG AAAATGTGAA AATTTAACAT	180
TTTTTTGTAC AAATTTTATA AATTATTGCC TTTTAAATAT CAATAGTTAA TCTCTTATCC	240
AGATCCCCCT TGTGTAAACT TTATCTTTAT AAGCTTCAAG GCCCCTATCC CATCTATTG	300
CAACAATTAG ATCACTTTGT TTTGTAAATA GTTCAAAATT CTTTCAATA ATTACGTTAT	360
CTATACTAAC GTTTAAATTT GGTTCATATA CTAAAATTTT TATACCGACA ATCAATAGTT	420
CATTAATTAT ACTTAAATA GCTGACTCTT TGTAATTATC TGAATTATAT TTCATCCCCA	480
ATTTATATAT TCCTACTATC TTTGGCTTTC GTTCCAATAT TTGTTTAACT ATGAAGTGT	540
TTCTATTGTG GTTTGAAATA TCAATCGCTT CTATCACTGG GGCATTTATT TCTATAAATT	600
CTTTTTTTAA TTGTTTAGTA TCTTTGGGAA GACAATATCC TCCAAATCCA AAAGAAGGAT	660
TATTATAAAA ATTTCCAATT CTTGGATCTA AACAAACACC TTTTATTACA ACTTCAGCAT	720
TTAAGCTTCT CCTCTCAGCA AAAGAATCTA GTTCATTAAA AAAGCAACAC GGAGAGCTAA	780
GAATGTGTTA GAAAAAGCT TAATTGCTTC TGCTTCAGTA GGAGAACTA ACATAACATT	840
TTTAATATTG GCAGTACTAT GAGTACTAAT CGAAAGGAAC AACTCTGCAA TTTTCTTCC	900
TTCAACTGTC TCATCTCCAA CAACTATGCG ACTGGATAT AAATTATCAT ATATAGAACA	960
ACCTTCTCTC AAAAATTCAG GGACAAAAAT GATATTTTTT GTATCAAACA GCCTTTTAA	1020
TTTGTTTGAA AAGCCGATCG GAACTGTTGA CTTTAAATA ATCTTTCCAT TAGGTTTAC	1080
CCTCAGAATC TTCGATACCG TTTGTTGAT TTCATATGTA TTAAACTAC CAATTTTCTC	1140
ATCATAATCT GTCGGAAGCG CAATAATATA ATAATCAATA TTATTTTAA TTTCAGAAAA	1200
TGTATCAAAA AAAGTAATAT TTAAGTTATT CTCGCAAAA AACTTCATAA GCTCTTCATT	1260
TTTAGATGGA AGAATGCCCT TTTTAAATT ATTTATTTT ACAGAACTA TATCATATGC	1320
AACAACTTTA TATTTAGATG CAAATAGTAA CGCGTAGGCC AGCCCAACAT GCCCAAACC	1380
AATTACTGCT ATATTCATAA AACTACTTCC TTATTTCTTA ATCCAAAATC TAATAGAA	1440
AGCTGCCCCA TTCCTTAAAT ACAACTCTTT AATATTGTTT AAAAGTTTTT CAACTGATTT	1500
CCAGATTATC AAAATCTGAG ATTTATAGCA CAATATTGAT GATATTCTAT CAATATAATT	1560
TTTTTCATCA AGTTCCTCTT GATACATTTT TAATTCTTTA GTTTTTCCCA TATACTAAC	1620
CATACTACTA TCACTTACAT ATGGGAAGTC CTCATAATAT ATTACTTTAT AACGCATAAA	1680
TTCAAGCGCC CTTCCAATAC TATTCACAAA AACATGAGCA ACATGGTCAC CAAGTGAAAG	1740
CGGACAATAT ACGACACATT TGTCGTCTAA ATGCATTAAC AGCTCTTTTA TGATATCATT	1800
CTTTAATGTG TCCTCATTTT TTAATTCACT ATAGATATGA CGGTATAGAA AATTGCCATT	1860

1023

TCTATCTTTC CTATAGAGAC ATTCATAGTA CGATAAGTGT CTAAAATCAC ATTGTAGACG	1920
TTCACAAGCT AACCTGTCTT CTTTCTTCCT TTCTTCAATC GGATATTTCC CAAGGTTACA	1980
CAACTTATGA AATTGCTTAG CAGAGGGCTG TAGCTGTTGG CTCAAAGGGT AACCAGAAAA	2040
TATAGTAATA ACAAGTACAA TTTCTCCTTC TGAAGTTAAT TTTGAAATAT AATCACCACA	2100
GGAAAAAATT GCGTCATCTA AATGTGGAGA TAAAAAGATA TACTTAGTAT TGTACTCAT	2160
AACCATTTCC TCTACAATTT ATCTAAAAAC TCACTAAGTG TCTGATTAAA TTCCACATCA	2220
TCAAAAAAAT TCACCTTATT CTTAATAATG AATATTTTCGT TAAATAAACA TATATATAAA	2280
TATTTCAATA TCCTTTCAAT ATCATCCTCT AAATTCTCCT CAATATTTTG TATCAGCCCA	2340
TTTACAATCT TATTAAAAAA GATAAGCTCT TTATCTCTAA AATTAAATAT TTTCATACAA	2400
CTGTTGTATC GAAAAATATA TAAATAAATT TTTACTAATG TTTGAATATT TAAACAACTA	2460
AATAAATGAG TTGTACCCGG GACACTATTT ATGTTATCAA GAACACTATC TTGAAACCTC	2520
AACTCACAGT TCTTTTTGTG AAATCTTTTT TTATCGTTTA CATCTGATAT TTTTTAGAC	2580
ATTTCAACAA TCTCAGACAT TTTATATGGA TATCTAGGAT GAATGCCAAA ACTATGCAAA	2640
ATGAACTGCA CCCCCAAAGT TAGACAGAAT AAATCTAACT TTTGGGGTGC AGTTCATAAG	2700
ATTGGGATAT TTTTTTTTAG CTAGAACTAG TAGAAATATA TAGTCAAATA ACAGATACCT	2760
TAAGGGTTTC TCATCTACAT AAAAAAATGA TACTTTTTTC TCTTCAGTAA TTACCTCATA	2820
AGCTTCACAA TAGAATCTCA TGTTTCCCTC CCCTATATTC TTAAATAAAA TCCTTTGGAA	2880
ATTGATATAT CTTAGTAAAA TATTGTTTAA GTTCCGGATG CGGAGCATGG GTAACAATAA	2940
TGACAGTCAA ATCCTCTCTA TCTAATATCT TACGTTCAAT CGCTAACGAA GTTCTCCTAT	3000
CGATAGCAGA AGTTCCCTCG TCAATTAATA CTATTTTCTT ATTTCTAATT AGCCCTCTAG	3060
CTAAAGTAAT TTTTGTTC TGCCCTCCTG ACAGTAATCT CCCATCATCA CCAACATAAT	3120
AATCTAAAAT GTTATTAGGA AAATCTTTTA CACTCAAACC AACTTGCTCT AAAGACTGTA	3180
GTATTTCTTC ATCAGTATAA TTTCTTTCCA ATAAAATATT ATCTCTAATC GTACCTTCAA	3240
ACAAATAAGC TTTTGTATCT ACATATAGAA CATTCGAAAC CATATTTAAA TAGGAGGTTT	3300
TTTTTATATC ATCCCCGAG AATCGCAATT CTCCACTATA ATCTCTCAA AAGCCATTCA	3360
ATAATTTTAA TAATGTAGAT TTCCCGCTTC CACTTTCACC TAAAATTAAA TACTTTTCAT	3420
TACGTTGAAA AAAAAAATTT AAGTTTTTTA ATATTTCTTT ATCTCCATAC TTATAGCAAA	3480
TATTTTTTGC TTCATATAAC GGAAAATCTC TATTCACCTC ATTTGGTTTC ATATCATTC	3540
TTTTATTGA CTCAATTGGA TTAATTGAAT ACAATTTTAA AAAAATAGGC TTCGTACCAA	3600

1024

TAATAGAGGA TAATTGACCT CCTAATTCAC CTAGCGCTGT AAAAATAACA CCTGTTAGTG	3660
CTCCTATTGC TTCAATAGTA CCAATTTTCA CTATTCCTTT TATTGCAAGA TAGCCTGTTA	3720
AAAAACGAG AGATATCTGA AAAAAATAT TGAGAAAGAA GCTAATAGCG CCTGCTAACG	3780
TTTCTACAGT TGTCTTTCTT TGTATAACCA TCTTTAATAA AATTCCTGCT TCTTTAATTT	3840
TCTTAGGCAA TACATATAAA AGATTCAAGG ACGCTAACAC ATCAAATCCA TTCAATATAG	3900
TCTCACTAGA TTTTAAAAAA GCTTCATTTT GGTAGTTAA ATTTAGACTA ACTTCTCGCA	3960
TTTTCGATGC AAAGATTTTT GGTACAAGTA GCATAATCAT TAATGAAAAC AAGGTGGCTA	4020
CAGTCAATGA CCAATGATAG TGATTAAGAG TCACAACTGC AAATATAGTA CCAGAAATTC	4080
CTTTTATTAC TAAAAAAGT TGTTTAAACG CCTGATCATT TAAAGTCTGA ACATCATTAT	4140
TTAGCCACGA AAGATATGTT CCTGATGATT TACTATGAAA TTCTTGATAG GTAGAGTTAG	4200
AGATGTCTGT GGCAACTCTA TTTCGAATCT CTAGATTAAA CTCTTGATC ACTTCAACCT	4260
GATAATTTTT CACTACCCAG TCAAGGAATA TTATCCACA CCAGACAATC ATTTGGTAGA	4320
TTGACAATTT CAAAAACCGC TCTAAATTC TCGCAATTAA TTCATTCAAC ACCAGAGCAT	4380
TAATAGTTGC TGCATAAATT AGCAATAATT GACCAGCAAC AATAAATATC GTTAATAAAC	4440
TAAATTTTTT TATATTGAT TTTATAATAG TATACACAAT AGTTTCTCAC TTTCTAAATT	4500
TTAATGAAC ATAGTTTTCA TATATACAAT AGAAAAACC AAAATGATAT AATAACATAT	4560
ATTTCAAAAA AGAAATTCGT TAAAAATTTT TTCTTCTCTT GCCTTCTTGA TTACTTTTAA	4620
AGCCTTGCAT TTGTCTCCTA TTAATAGTAA CCGCTTTATG TTTAAAGAAT AATATTTCTT	4680
TGTAACCAAT ATTCTCTCGT TGAAACTCAA TAAATTAAAA TATTTCTAC AGTAATTATA	4740
ATATTCTTCA TCTGCATTAA TTGTTTTTTG TGTCACCTCA GTGATACCGT TTTCTTTACT	4800
GTGAGCGTAG TAATTCACCA AGAATCTCTG CACTATATCA ATTTGGTATC CTTGAACAAG	4860
TAGTTTTAAT AAAACAACAC CGTCTGATG TGAATCTATT TTCTCAAAAC CATTAATTAA	4920
TTCTAGCACC TCTTTTTTAC ACAACCAAAA TGACGTACCT GCTATATTGT GAACCATTTG	4980
AACAAACAAG GGATTTCCAA CAAAATCGGT CTTCTCCTCT TCTCGTGAC CATTTGGATA	5040
AATTATTATT CCATAACTAC AACTAAAGC TAAATCTTC ATTCTACTCT TTTTAAACA	5100
AGCCATCAAC TTTAAAATTC GATCTGGCAT ATATTCATCA TCATCGTCTA AAAATGATAT	5160
ATACTTACCT CTAGAATTTT TGATACCTAT GTTTCTGGCA TTAGTTGCAC CTAAATCTTC	5220
ATTACTTAAA ATTAACCTAA TTCTATGATT GGTATAGCCA AATTGATGGA TAATTTTATT	5280
TCTTAAATTT ACATTACTAT AATTATCATC AATAATTATA ACTTCGATAT TTTTATAACT	5340
TTGATGTAAA CAACTTTTCA CAGCTCTAAT CAGAGATTCA TACCTATTAT GTGTGGTAT	5400



1025

TATAATACTT ACTAATTCTT GATCTATATT CCTATCCATG ACTACTCTTC TCTAATAATT	5460
CATCATATAC TCTCATGGTT TCTACAAACA TTTTGTGCAC AGAAAAATGT TTTCTTATTT	5520
TTGATTTACT ATTCTCACCT ATATATTTCA AATACTCAGA ATCATTGAGT AAAAAATTAG	5580
CACAAGCACA CACTCCCTCA ACATCTTCCT TCTCAAATAA AAATCCATCA ACCCTATGTT	5640
CAATAATTC ACTTAACCCG CCAACATTAC TAGCTAAAAC CGGAGTTCCT TGTGACATTG	5700
ACTCTAAAAC ACACATAGGT ATTCCTTCTG TATCAGAAGG AATATACAAT AAATCCGATA	5760
TTTGGTAAAC TATAGTAGCT GGATAGATTT CACCAAGTAA CCTGAAATTA TCTCTACATT	5820
TCAAATGGCA AATTTTTTCT TTCAAAGCAG CCCACATACT ACCATTTCCTA GCCATAATAA	5880
AAATCACATC TTCTCTGACT AAAAATAATT TTTCTGCAAA TTCAAGGAAT CTATCCGGCC	5940
TTTTTCTGG ATCCAACCTT CCAACATAAC AAATGATTTT TGTATTATTG GAATACAAAA	6000
TTCTTTTTTA AAGTCTTGAA CACCTACTAC ATCTAAATCG CTATTTGATA CATTAATTCC	6060
GTTATTTATT GCAACTATCT TCTTATTTTT TATTATACTC TCCAATCTTT TTTTCATAG	6120
TTTCAGATAC ACAAATAAAA GCATCTCCCA TAGAATATGT CCAAAAATCA AAATAAGTCA	6180
AGAATTTCTT TTTTAAGTTA TATTCAACCC ATCCATGGCA TGTATCACT GTCTTAACCT	6240
TTCCAAATCC ATTCTTGTC AATTTTTTTA ACATATATAA AAAATAATTA GTTGAGTAGC	6300
CATGACAGTG TATAAGTTGG ATTTTAAATA ATTTTAAAT ATTTTAAACG TGTAAGGCAG	6360
TTTCAAATTT ATTTGAACAT TGAGTACAAT CAACATAGGC AATATCTAAA TTTTATAAT	6420
CATCAATAAC CTTTGAATCT CTAGATACAA TTATCAAAAT AGGGAATAGA GACA	6474

(2) INFORMATION FOR SEQ ID NO: 156:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4792 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

TATTTAACGA TTTTTCAT GTCATTCCT CCAAATAGA ATACCTTATA ATCTTAACAG	60
AAAAAGAGCA TTTACGCCAT TATATGATAT CTATCTCTGT GATAAGTTTT TTTATGGGT	120
AATTTAAAG ACCAAACGCA AGATGGCAAT CAAGACCACT CCAAAGAGAA CTGTTCCGAC	180
TAGATTGCGG TAGCGAAAGG CTACCCAAGC TGTGGAAAG ACGGCTAAGA AGTCCAGTCA	240
TTTGATTTGA GGAAGACTGC CAACCTTACC TGTCACTACG CTTGAAAGAA TCAGGGCAAA	300

1026

GATAATGGAA ACAGGCAAAA ACTTCAAAAA ACGCTCAACA ATCGCAGGCA GGCCCTTATA	360
CTTGACCAAG ATGAAGGGAA TCATACGGGG AATCCAAGTC ACCAAGCCAG AGAAAATAAC	420
TGCTAATAAA AGATACTTAC TGACCATCTA AAACCACCCC CATGCTACAA CCAAGTAGCG	480
TCGCAAACAG AACAGCTAGT GACTGAGACA TCACTGTCAA GAGCAAAAAG AAGGACACCG	540
CAACAACCTGC TAGGATAATG AGCAGATTGC GGACAGGAAT CCGTCTTTGC ATAATCTGAA	600
ATTGCGAAGC AAAATACCAA TAAACATCCC AACCAGGGCA AAATCCAAGC CAAAGATTTT	660
TGGATTTGGT AGCAGGCCAC CCAGAGCCGT TCCGACTACT GTCCCCACAA ACCAAGCCAC	720
ATAGCTGTTA AGATTGTTTC CGTGCATCCA CATAGGATTT ACCTTGCTCTG TATGGGCCAA	780
TTCACCCATC AAAACGCCAT AGGTCTCATC TGTCAAGATA CTAGACATAC CGATATTGTA	840
CCAAAGACTG GTATGACGGA AATAAGTCGA TGCCTGTAAA CTCAACAAAA AGAGACGCAA	900
GTTGATTAGA AAAACCGTCA TAGCAATAGC TGCCACAGGA GCTTGAACCA CAATCAGTGC	960
CAACATGGCA AACTGGGCAC TCCCAGCATA AACAAAGAGA CTCATCAAGC CCATCTCAAC	1020
AGGTGTCACA TAGGGCGCAC CGATAATTCC ACAGGCCAGG CCGATACTGA CATAGCCAAG	1080
AGCCGTTGGC ATGGCTGCCT GCGCCCCCTC CTAATACTCT TTTTCTTTCA TCTTTCTCCT	1140
CATATTGTCT TAATAATACT CAATGAAAAT CAAAGAGCAA ACTAGGAAAC TAGCCGCAGG	1200
TTGCTCAAAA CACTGTTTTG AGGTGTCAGA TAGAACTGAT GAAGTCAGCT CAAAACACTG	1260
TTTTGAGGTT GTGGATAGAA CTGACGAAGT CAGCTCAAAA CACCGTTTTG AGGTGTGGA	1320
TAGAACTGAC GAAGTCAGTA ACCATACCTA CGGCAAAGTG AAGCTGACGT GGTTTGAAGA	1380
GAGTTTCGAA GAGTACAAGT AGGCTGAAAA GAATCCAACC ACAGCATGGA CTATTATATA	1440
GCAGATTGAA ATAAGATGAG AACAAATCGA TTGGGAAAGT AAAATTAATT TCTATAAATG	1500
TTTTAGCAAT TGTTTCGTAC TATTTTAGAT TCAGTCTATT ATAACACATT CAGAAAAGAG	1560
AAAAAAGTCT GTTGATTTTG ACCATCATAA AAAGACTGGC AATCCAGTCT CAAACATATA	1620
TTATAGAAAT TCTCCACTAA ATACTTTCAC GAATATTCAG AAGCATAACA AAGGCAACTA	1680
GAAGAAATAG CAATAAAACA AAGCTAACTG CCAGAGTTCC AAAGCTAGTA GCAATGGTTA	1740
CCAAAGCTAT TGTAATAAG CTAGGTAAAA CAACCGTAAT GGCACCGATA GAGGATTGAA	1800
CTGCTCCCAT TGACTCCTCA GGTATTTGTT TAAAAACGAG TTCTTGCAAT CTAGGAGAGA	1860
GAACACCTGC GAAAAAGGCA TCCAAGGTAC TAAAGATGAG AATCCAGTCA AAACGAACTG	1920
TGGCAAATCC TACTAGAAGA AGCAACTGGA TGACAAGTGA GGCATAGAGA GCTGTTTTTA	1980
TGGAAATGGT ATGTTGCAGA TAGCCACTTA CAAGGCTTCC GACAATCAGG GCTGATAATT	2040
CTAGTGTGGC TAACAAGGCA AGAGATTGAC CAGTTTGTA AATCAAAAAG GGCTGGTTCC	2100

1027

TTAAAAATAG AGTGGAATA GGAACCGTAA CATTTATCAC TGCTTGAATA GTAGAGATAA	2160
TAAACAAAAC CAAGAGCACC TTATTCATAT TCCATATCAA TTTCGATGAT TGGAGCAAAT	2220
GCTGGCAAAA GGATTTTACA GAGAGTCCTT CTTGATAGCT AATCGTTTTT TCTACTTTCA	2280
AGAGGTCAGT TTTTATGAAG AGGATACCTA AAAATGCGAT TAAAAAGGTA AGAGCGTTCA	2340
GTAAGGAAAT AAAGTGGATG GATAGAATGC CTAGTAAGAC TCCTCCTAGG ATATTACTGA	2400
<hr/>	
TTGTTTTTAC TAAACTAACA GTTGACTGTT TAAAGCCAAT AGCTTCTGCC AGATGGTCTT	2460
GGCCAAATAT TCTAATGAAA ATCGGAGTGA GCATGGCGCC TGAAAAATAA CTCAATGTGT	2520
CAGACAAGAG GTTAATCAGA CAAATAAATG CTAAGTAGCA CAAGGAGAAA GACTGCCCTG	2580
AAAGTGATAA AGACACTATA GAGTAAAGCA AAAATTTTGC AAAACTAATG ACTGTGTATT	2640
TCAAGACACG ATGATGTTGA AAATCCGCCA AAAGTCCAG AAAGATTTGT AGAAGTTGGG	2700
GCAGGGTTTC TGAAATCGTG ATGAGTAAAA TCGCCAAAGG GGCAAAAGAT GCATCTGCCA	2760
CATAATTCAG GAAGGCCAGA TAAAAATCG TATCCCAAG CGTTGAAATC CACTGGTTGA	2820
TAGTTAATTG CCTAAATCT CTATTTTGAA GAAATACTTT CATCACAAT CTTCTTTAAG	2880
TTCAATGGG AATCTTTCCC CAAGGATAGA CCGCGATACT ACTAACAACC AAAATTACAG	2940
TAACATCAAA AGCTGACCAA TGCCATTGTA GACTATATGC AGTCCAATAG GCCAATAAAT	3000
TGACTTTGTC ATTCTAAATA AGACTGCAAA TATAAGACCT CCACCCATAT AGAAGACAAA	3060
GTCTGTCAAG ACCCAACCGT GATTACTAAT GTGCGAGACC CCAAATAAAA CAGCGGAACC	3120
AAGTACATCT AGCCCCCATT TCTTTCCTTT TTCCAGAGCA GTCATCACTA ATCCACGATA	3180
AATCATGTCT TCAAAAATGG GACCTGCAAT CACAGGATAA AAAAAATACA TCAAAAATGC	3240
TGTAGCCCCT GTAAAAGTCG GAGCAGCATG TTGATAAGAA ATTTCAATTC GAGTAGGTGG	3300
GAAAAGAAAA AAGGTAACGA AATTCCAAAC AACAAAAGCA AGCAGAGCTA GGAAGGAATA	3360
GAAAAGATAG GATCCTTTAA ACTTTCTACT ATTGATTTTC TGCCATTTCC CCGACCAAAT	3420
CATAGCAATA AGAGCAAATA AAACCACAAG AAAATTCAAC ATCATATCCG ACAGATAATA	3480
GGCAAAGTCA GATAGCCCAG TAACAAGGTC GCTGCGTAAA ACTAGAACAC TGAAGTTCTG	3540
GTCAGCAATA ACTAGTAGAA AAAGTATAAT AAAGTAGCGG TGTGAGATTA TCTTTTTCAT	3600
ATATCACCTT TCTAATATCC AAATACCAAT AAAGTAACAA TGAGTAAGAA ACTATTCCAT	3660
GAAGCATGCA GAGCTATAGC CCAATAGATG GATCGGGTGT AGCGAAACAT CATACAAAAT	3720
ATCAAGCCCA TTCCAAAATA CTTTATGAAA TCTGTCGTTA TCCAACCATA CTGCAAAACA	3780
TGCATAGCGC CAAATATGGC AGCGGAAACA AGAACATCAA GATAGTATCT CTTAACTTTA	3840

1028

GATAAACTTG TCATCAAAAG ACCACGACAA ACAACCTCTT CTGATACAGG TGGGATAATA	3900
CTAGTATAAA GTATTGCGGT AACAAAATAG CTAATTCCTG TTAAATTGGT GGCTACTTCT	3960
ACGACTGTAC TTCCATTCTG GGTACGAGGA AAGATATAGG TTGTTAGATT TGGCCACACG	4020
AACAATAAGA AAAAAGAAAG AAGGAAAACA CCCAGGTAAG ACCAACGAAA CTGGAAACGA	4080
CCCACTCTT TCCAATGTTT ACTTTTGACA AAAGCAATTG TAGCTATAGT TCCCAGAATA	4140
AGTACCAATA AAACCTGGAA CACATAGTAC ATATTATCAG ACAAAGCAAC CATAAAATCT	4200
AAGTCTGATG TGACATTAAA AATGAGGTAA TAAGTCAAAA TCAACAAGCC AGTTGCTAGG	4260
TGAAATTTCA CTTCTTTCAT TTTCTTCATC CTATTATCTC CTATAAGAGC CTATCTTCTA	4320
CGGCGGCCAA ACAATCCATC TGCTAAATCT ATAGTCCAAT CAAAAGCTCC ACGATTAGGA	4380
CTCATCCCTT GATTGCCCCA ACCAGGGTAA ATTCCTGGGA CGCCCCAACC AGATATACCA	4440
CTTCTTCCAC CACCTCCCAT AGAATTTACG AGGTGCTCTC CTCTAACATC TTGCAACTCA	4500
GCTTCTGTCA ATTCCATTGT TTCTGCAAAT TGTAAATTTA ACATCTTTTA CACTCCTTCA	4560
ATTATCTTCA TTTGTAAACC ACTTCTGCGA CCTAGGATTG GCTTCAAGTG CTTTACAAGT	4620
ACAGTATAAC ACGAACATTG GCTTATTTTA GAAATCGCA TATTTGATAT TTTTCTTAT	4680
AGAAATTTCA GATTGCGAT TTTGGTGAAT TTGATTACTT CTCTGGTATA ATAAAGTTAC	4740
TACTAATGAG GAGTGGAGAA ATATGAAGAA ACAAATTTTA ACATTATTGA AA	4792

## (2) INFORMATION FOR SEQ ID NO: 157:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2156 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

CCGTTCTCGG CGACGGCCAT CTGATGAAGC TATTTATGAG GGAACTGGC AAGCTGGAGA	60
GTCAGAGTAT CTAGTCTTTC ACCGATTGCT GTGGCAGCAG ATGTGCAGGG AAAAGGAGTT	120
GCTCAAACCT TCTTAGAGGG CTTGATTGAA GGTTTTGATT ATCTTGATTT TCGCTCAGAT	180
ACGCATGCTG AAAACAAGGT TATGCAACAT ATTTTGTAAA AACTTGTTT TAAACAAGTC	240
GGTAAGATGC CAGTAGATGG CGAACGCTTG GCCTATCAAG AATTAAAGAA ATAATGCAAA	300
AGAAGTATGT AAAATCCTC TACTCCTCAC CAATTGGTAT TCTATCACTT GTAGCTGATG	360
ACCATTTATT GTATGGAATT TGGGTTGAGG AGCAGAAGCA TTTTGAGAGG GGAAGTAGG	420
ATGAAACGAT AGAAGAAGTT GTTAGTCATC CTATTTTAGA CCCAGTTATT GCTTGCTTAG	480

1029

ATGATTACTT TAAAGGCAAG CCTCAGGATT TATCCAACCTT GCTCTTGGCG CCAATCGGAA	540
CGAATTTTGA AAAGAGAGTT TGGGACTATT TACAGGGCAT TCCTTATGGT CAGACACTGA	600
CCTATGGACA AATTGCTCAA GACCTGCAAG TGGCTTCTGC TCAAGCAATT GGTGGAGCAG	660
TGGGACGCAA TCCTTGGTCT ATCCTAGTAC CTTGTCATCG TGTGTTGGGA GCAGGCAAGC	720
GTCTGACAGG TTATGCTGCA GGAGTGGAAA AGAAAGCTTG GCTCTTGGAG CATGAAGGAG	780
TAGATTTTAA AGATAGAAGC AATAGAAGGA GAAGCACATG TTAGAATTTA TCGAATACCC	840
CAAATGTTCA ACTTGTA AAA AAGCAAAACA AGAATTAAAT CAATTAGGTG TGGACTATAA	900
AGCCGTCCAT ATCGTGAAG AACACCTAG CCAAGAAGTC ATTTTGAATT GGCTAGAAAC	960
CTCAGGATTT GAATTGAAGC AATTTTCAA CACCAGTGGT ATCAAATACC GTGAATTAGG	1020
GCTAAAAGAT AAGGTAGGAA GTTTGTCAA CCAAGAAGCG GCTGAGTTGC TAGCAAGTGA	1080
CGGTATGTTG TTA AACGGC CCATTTTAGT AGAAAATGGA ACTGTTAAGC AAATCGGTTA	1140
TCGAAAATCT TATGAGGAAC TGGGACTGAA ATAGTTTTTA TCTATCTCTT TGATAGATAA	1200
AATATATAAC TTCCCTGTTT CAAAGTATGA TAACTAGTA GGTAGACAAA GTCTGTATCT	1260
GACCGTAGCA AATAATTTCA TTGACGGCAG AAGCATGGTA GCATGAATCA TTATCAGAAG	1320
AGGATGTTTT TATGAATGTT ACAACGATTT TAGCATCAGA TTGGTACCAA AACTTGATGC	1380
AATTGATTCC GGATGGCAAG CTGTTTAGCC TACGTTCCGT CTTTGATGGA ATCCCTAGAA	1440
TTGTCCAACA ACTTCCAACA ACAATTATGT TGACAATTGG TGGTGCCCTT TTTGGCTTGG	1500
TTTTGGCGCT TCTTTTGCC ATGTGGAAGA TCAATCGTGT CAAGATTTTA TATCCCTTGC	1560
AGGCCTTCTT TGTTAGTTTC TTAAGGGA CACCGATTTT GGTGCAACTC ATGTTGACCT	1620
ACTACGGAAT CCCTTTGGCT TTGAAAGCCC TCAATCAGCA ATGGGGAAC TGTCTCAATA	1680
TCAATGCGAT TCCAGCTGCA GCTTTTGCGA TTGTCGCCCTT TGCCTTTAAT GAGGCAGCTT	1740
ATGCTAGTGA AACCATTCTG GCAGCCATTC TCTCAGTTAA TCCTGGTGAG ATTGAGGCGG	1800
CACGCAGTCT GGTATGACC CGAGCGCAAG TTTATCGACG AGTGATTATT CCTAATGCAG	1860
CGGTGGTAGC TACTCCAACC TTGATTAATT CCCTCATCGG TTTGACCAAG GGAACATCTC	1920
TAGCTTTTAG TGCGGGTGT GTGGAAGTCT TTGCCCAAGC TCAGATTCTA GGTGGAGCTG	1980
ATTATCGCTA TTTGAACGC TTCATCTCCG TTGCCCTTGT TTATTGGGTA GTCAATATCG	2040
GAATTGAAAG CCTCGGTCGT TTCATCGAGA GAAAAATGGC TATTTCTGCA CCTGATACAG	2100
TGCAACAGAT GTGAAAGGAG ACCTTCGTTA ATGATTAAGA TTTCGAATTT AAGCAA	2156

(2) INFORMATION FOR SEQ ID NO: 158:

1030

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 3140 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

GTATCTCTAC ACATGTCTTC AATCGATTTT GTTGTCTCTCC AATTTAATTC CTTATATGCT	60
TTGTCTGCAT TTGCATAACA AGTTGCAACG TCTCCTGAAC GTCTTGGAAC TATTTTATAA	120
GGAATAGGGA TCTTATTAAC ACTTTCAAAT GTATTTACAA GTTGTAATAC ACTAGTGCCT	180
TCTCCCAGC CTAGGTTATA GATATAAACA TCTGTTTTTT CAGATACTTT TTCTAAAGCT	240
TTTATATGTC CTATTGCTAA ATCTACTACA TGGATATAAT CACGCACACC AGTACCATCA	300
AGCGTATCAT AATCATTTCC GAACACACTT AGCTCTGATA GCTTACCTAC CGCTACTTGT	360
GCAATATAAG GCATCAAGTT GTTAGGAATT CCTGAGGGAT CTTCCCCAAT CAAACCAGAC	420
TCATGAGCAC CAATTGGATT GAAATAACGA AGCAACGCAA TACTCCATTC TGAATCTGCC	480
ACATGAACAT CTTTTAAAT TTGCTCAAGC ATCACTTTTCG TATACCCATA AGGATTGTGC	540
GCACTTGTTT GCATCGTCTC AATTAGAGGT GACTGATTGT TAATTCCATA TACAGTCGCA	600
CTTGAAGAAA AGACAATCTT TTAAACATTA AATTCTGACA TCACTTCAAC AAGTGCCAAT	660
GTACTCATAA TATTATTTTT GTAGTACATC ACAGGCTTTT GCACGGATTTC TCCGACAGCT	720
TTATAACCTG CAAAATGAAT TGCAGCATCA ATCGATTCTT GTTCAAATAC CTTTCTCAAT	780
GCTTGTTTAT CACAAACATC TAATTCGTAA AACACGGGAC GTATTCCTGT AATTGCTTCA	840
ATACGGTCTA GCACCAAGAT GCTAGAGTTC GAAAGGTTGT CGACAATGAT AACTTCCTTT	900
CCTAAATTTA GTAATTCTAC TACGGTATGG CTACCAATAT AACCAGCTCC GCCTGTTACC	960
AATATTGCCA TCTGGGTTTC CTCCTAATTA ATTCCAACCG ACTTAACAAA TCTCATAAAC	1020
GCTTCATGCC CAGACGGTGT ATTCTTATAA ACTCCTGCAT CTTCCAGAAC TCTCGCAAC	1080
ACTTGTCTG CTTCGTGTG AACTACGCTA TTAACCTCTT CTTTATTAAT GCGAGGATAT	1140
TTTTCTTTCA ATTGGTCGGC CCATTCTAAA TGATAATCCG CAATTGCATT ATCCTCTCCT	1200
AAAAGATATT TTCCAACCTC TTCTAACTCT GGTTCAAAC GAGGTGGTAA TATCGCAAGT	1260
CCCATCACTT CGATTAACCC GATATTTTCC TTTTAAATAT GTTGTAATC TTGATGAGGA	1320
TGGAAAACAC CATCTGGGTA TTGTTCACTA GTATGATTAT CTCTTAGAAC AATATCTAAT	1380
TCGTATCTCC CGTCCACTTT ACGAGCAATA GGAGTCACCG TATGGTGTGG GACATCTTCA	1440
GTCATAGCAA TGATGTCTAC TTCTAAATCT GAATATTCTC TCCACTTATT TAGAATTTTA	1500

1031

GTAGCTAAAT CTAACAAGCG ATTTTATTT TCACTTTGTA ACCTAATTAC TGACATTGGC	1560
CATTTTACAA TACCAGCATT AACATCCTCA AAGTCTTTAA AACAAAATTC ACTCTCAAAT	1620
TTTGCTTTTT CCATTGGGAA AATATGTTTC CCTCCCTGGT AGTGGTTATG ACTAAGAATG	1680
GAGCCTCCTG AGATAGGAAG ATCAGAATTT GAACCAGCAA AATATCCTGG CAAAATATCA	1740
ACAATCTCCA ATAATTGTTC AAATGTTTGA GAGGTAATAG CCATTGGTAC ATGTTGACTA	1800
TTCAAAAATA TCGCATGCTC ATTAAAGTAT GAGTAGGGAG AATACTGGAA TCCCCATACT	1860
TCGTCAACCA GTTCAACCG AATAATTCTA TGATTGGAAC GTGCTGGATA ATTTATTCCG	1920
CCCTGATATC CTTCAATTTT CATACATAGT AAACATTTGG GATAATTAGT TGCTTTTACT	1980
AAATTTTCAG CAGCAATTGT TTTTGGATCT TTTTCGGGTT TTGACAAATT TATCGTAATC	2040
TCTAGCTCTC CGTATTTAGT TGATGCTCGA AACTCAATAT TCTTAGCAAT AGCAGAAGTT	2100
TTAATATAAT CACTATCTTT ACTTAACTTA TAAACTCTT CAACTGCTTC TTGAGGTGAT	2160
ATATCATATG AACTCCAAAA AATATCATTT AATCGACTAG GTAAAGGAAC TATGAAATTC	2220
ATTAACTCTG CTCCTAAACA TTCCTTTTCC TCGATTAAAT CTTTAATTTT ACCGTTTTTT	2280
AAGGCGATTT CCACTAAGTA ATCTTTTATT TGTTTCAGGT CATTTTCATC GGAAATGCCA	2340
TCAATCCCTT CCTCACCTAT TAACGCTAGT ACTCTATTTT TCACATATAT TTTGTCAATT	2400
TCATTATACA TTCCGTATTC AATTACTCTA TCAACAAAAT TATCAATAAT TGTTTTCATA	2460
TATTTTCTTT TCTAATTTAT GTTCCCATAT TTTCTATACA TTATCCATTT ATAAATTGCT	2520
TGCGTAGTAT GAGCAATTTT ATCAAGGTGA TGAATAATAT CTAAAGCACT AATTACTTCA	2580
GAAACGTTCC CATCATCTTC AAATATGTAA TTCATTATTT TCTTTTCCAT ATTTATACTA	2640
AGCTCTTCTA TCTCATCTCG TTTTGTGATA ACAACCATAT CTAAACATCC AGATTGTTCC	2700
TCTCTATAAC AAGATATAGC CCTATTCATA TGCAGTCCGA TAACTTCATG AAGTATTTTT	2760
ATTTTGTAAA TAATTTTCTT CAAAATTTCA TTATTTTGAA GAATCTGTAG ATTTTTTAAA	2820
ATTTCAACAA TTCTATCCCC AATACGTTCA ATGTCAGTTG ATATTTTAT TACACTAATA	2880
ATTCTTCTTA AGTCATATGA AACAGGATGT TGTAAACAAA TTAATCATA TCCTTTTTTA	2940
TCAATATTTA GAACTGACTC ATTTATGATT AAATCTTCTT TAATCAATTC TACTCGTTCT	3000
TCATTTGATA AATATTCAAA TAACTTCTCA TATTTATCAA GCACAGATAC CCAAATGGTC	3060
TCTAAATTAT TTGATAATTC TATAATTTCA TTTTCTAAAT ATAACCTTAA CATTTAGGTA	3120
CCTCTTCTTA ACAAAGTTCC	3140

(2) INFORMATION FOR SEQ ID NO: 159:

1032

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9048 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

CCGGATGATT TCCTGGTCAG ATAGGGGGAA AGTGACTTCC TCAGCAATCG CGCGTAGAGT	60
AGGATTCCCT TCACGGATAA TATCGTTCAT ATCAATTAAG TGAGCAGCTT TTGTAATACG	120
TTCTATTGCA GACATTTTCT CTCCTTATAT TATGTTTAGT GCAGTTAGCT ACTGCCAAAG	180
CCCAAGTGGT AACTTTGGAA TAAGCCACTG TGGATTAGTT CATTTTCTTT CATTACCTCT	240
ACATGATATC ACAAATGAC AAGAATTGAA AGCATTATGG CATTTAGGAT TTATAGAAAA	300
TAGATAGGAA GTTCAATTCA ATTGTGAAAG AAATACTTAT CTGTGATATA ATAAAAAGAA	360
AAGGCTTGCA TAAGAAAGTA GGGAGAACGA AGATACAAAG AAGACAAAAT CGAAATCAGG	420
GTGGTTTAGC TTTTCGTTTT ATGAAGGGCT TGGTAAACTT TTTAGGAGTT ATCGCAAGTG	480
GAGCAATAAG GGATTGTGG CGATACTCTT GCTAGCAGTT GGTATTATCA TGGGCTTGGT	540
CTTGTGTTTT GAAAGCTTCC AAGGAATCCC TTGACTAGTC AAAAACGAGA TACTATTTCT	600
CAAGAGGGGA CTAAGCAAAA GTCTCAGGAG TAGGAAGAGG AAAAACTGC CAGAATTATG	660
GCCCACGGGG ATTTGCTCTA CCACGATGGA CTTTTCTTTT CAGCTAAAAA AGAAGACGGT	720
ACCTATGACT TTCATGAAAA TTTTGAGTAT GTGACTCCTT GGCTCAAGCA AGGGGACTAA	780
GCAGCAGATT TAGCTATTGG TGATTTTGAA GGAACCATTA ATAAGGATCA TTATTTAGCG	840
GGTTATCTTC TCTTTAATGC TCCTGTTGAA GTTATGGATG CTATTAAGGA GGCAGGTTAT	900
CATGTGCTGG ATTTAGCTCA TAATCATATT TTGGATTTCG AAATTGAGGG AGTTATTICA	960
ACGGCCGATA TTATTGAGAA AGCTGGAATC ACTCCAATCG GAGTTTATAC GCACGAACCA	1020
CGTGATCAGG CTCGCTGGT CATTAAGGAA GTGAATGGTA TCAAGGTTGC ATTGTTAGCC	1080
TATTCCTATG GTTCAATGG AATTGAGCAG TATATTTCTC AGGAAGACTA TAATCGTTAT	1140
CTTTCAGATT TAAACGAAGA TAAGATGAAG GTTGAAATTG AACGGGCAGA GAAGGAAGCA	1200
GATATCACCA TTATCATGCT TCAGATGGGT GTTGAGTATC GATTGGAACC AACTGAAGAA	1260
CAAAAAGCTC TTTATCACAA GATGATCGAT TTGGGAGCGG ATATTATCTT TGGAGGGCAT	1320
CCTCACGTTG TTGAACCATC TGAAACGGTT GAAAAAGATG GAGATAAGAA ACTCATTATC	1380
TATTAAATGG GGAACCTCAT TTCCAATCAA CGAATTGAAT CTATGGGAGA TGAAGAGAAT	1440
GCTAAGTGA CTGAACGTGG TGTTCATG GATGTCACCA TCAAGAAGAA GGATGGAAAA	1500



1033

ACAACTATCG GAACAGCTAA AGCTCATCCT ACTTGGGTCA ATCGAACACC AAAGGGAACC	1560
TTTTCAACCAG AAGGATATCC CTTGTATCAT TACCAAACTT ATATTTTGA AGATTTTATA	1620
GAGGATGGCA GTCATCGTGA CCAGTTAGAT GAAGCGACTA AGGAACGAAT TGATACAGCC	1680
TATAAAGAAA TGAATGAACA TGTGGGATTG AAGTGGTATT AGCTTGAATC CAGAGGAAAAG	1740
TAAATGATGA TTAAGGTAAT TGCGACAGAT ATGGATGGGA CCTTGCTGGA TGCTAGAGGT	1800
CAGCTTGATC TCCCACGATT GGAAAAGATT TTAGATCAGT TGGATCAAAG GGGCATTTCGT	1860
TTTGTCAATTG CGACGGGCAA TGAAATTCAC CGCATGAGAC AACTACTGAG TCCCTTGGTG	1920
GATCGAGTGG TTCTGGTTGT TGCTAATGGC GCTCGTATTT TTGAAAACAA TGAATTGATT	1980
CAGGCTCAGA CATGGGATGA CGCCATTGTC AACAAGGCTT TGACTCATTT CAAGGGTCGA	2040
GCGTGTCCAGG ACCAGTTTGT TGTAACGGGG ATGAAGGGTG ATTTTGTCAA GGAAGGTACG	2100
ATTTTACAG ATCTTGAAAG TTTTATGACT CCAGAAATGA TTGAAAAATT CTACCAACGG	2160
ATGCAATTTG TGGATGAATT AACATCTGAC CTCTTTGGTG GTGTGCTCAA GATGAGCATG	2220
GTGTTGGTG AGGAACGTTT GAGTTCGGTT TTGGAAGAAA TCAATGCTCT CTTTGATGGC	2280
CGTGTCCGAG CTGTATCCAG TGGCTATGGT TGCATTGATA TCCTCCAAGC TGGGATTCAT	2340
AAAGCATGGG GCTTGGAGGA ATTACTCAAG CGCTGGGACT TGAAATCCCA AGAAATCATG	2400
GCTTTTGGTG ATAGTGA AAA TGATGTTGAA ATGCTTGAAA TGGCTGGAAT TGCCTATCGG	2460
ATGGAAAATG CTGATGAGAA AGCCAAAGCT GTGGCGACTG CTCTAGCACC AGCCAAACAGC	2520
CAAGGAGGAG TTTATCAAGT CTTGGAAAAC TGGTTAGAAA AAGGAGAAATG AAGTGGCAGT	2580
ACAGTTATTA GAAAATTGGC TCCTAAAGGA ACAAGAAAAA ATTCAAATA AGTATCGTCA	2640
CCTAAATCAC ATTTCTGTTG TAGAACCAAA CATTCTTTT ATTGGGGATT CCATTGTCGA	2700
GTATTATCCT CTACAGGAGC TATTTGGGAC TTCAAAGACG ATGTGCAATC GAGGAATTTCG	2760
TGGCTATCAG ACAGGACTGT TACTAGAGAA CCTTGATGCT CATCTATATG GTGGAGCAGT	2820
AGATAAAATT TTTCTTCTGA TTGGGACAAA TGATATCGGA AAGGATGTTT CTGTGAATGA	2880
GGCTCTCAAT AATCTCGAAG CTATCATTCA ATCCGTTGCT CGCGATTATC CATTGACAGA	2940
GATTAAATTG CTTTCCATTT TGCCGTGTCAA TGAGAGAGAG GAGTACCAGC AGGCAGTCTA	3000
TATCCGCTCG AATGAAAAA TTCAGAACTG GAATCAAGCC TATCAAGAGC TTGCATCTGC	3060
CTATATGCAG GTGGAATTTG TGCCAGTATT TGATTGTTTG ACAGACCAAG CAGGCCAACT	3120
CAAAAAAGAA TATACAACTG ATGGACTGCA CCTCAGTATT GCTGGTTATC AGGCTTTGTC	3180
AAAATCCTTG AAAGACTATC TTTACTAAAT AGCTAAATAA TGTTAAATTT GAGCATAATA	3240

1034

TCTTGTA AAAA	AATTCTAAAA	TCCTTTAAAA	TAAAAAGTGA	CGGAGGAATT	TATGAATGTA	3300
AATCAGATTG	TACGGATTAT	TCCTACTTTA	AAAGCTAATA	ATAGAAAATT	AAATGAAACA	3360
TTTTATATTG	AAACCCTTGG	AATGAAGGCC	TTGTTAGAAG	AATCGGCCTT	TCTGTCACTA	3420
GGTGACCAAA	CGGGTCTTGA	AAAGCTGGTT	TTAGAAGAAG	CTCCCAGTAT	GCGTACTCGT	3480
AAGGTAGAGG	GAAGAAAAAA	ACTAGCTAGA	TTGATTGTCA	AGGTGGAAAA	TCCCTTAGAA	3540
ATTGAAGGAA	TCTTATCTAA	AACAGATTCT	ATTCAATCGT	TATATAAAGG	TCAAAATGGC	3600
TACGCTTTTG	AAATTTTCTC	ACCAGAAGAT	GATTTGATTT	TGATTCATGC	GGAAGATGAC	3660
ATAGCAAGTC	TAGTAGAAGT	AGGAGAAAAG	CCTGAATTTT	AAACAGATTT	GGCATCAATT	3720
TCTTTAAGTA	AATTTGAGAT	TTCTATGGAA	TTACATCTCC	CAACTGATAT	CGAAAGTTTC	3780
TTGGAATCAT	CTGAAATGGG	GGCATCCCTT	GATTTTATTC	CAGCTCAGGG	GCAGGATTTG	3840
ACTGTGGACA	ATACGGTTAC	CTGGGACTTA	TCTATGCTCA	AGTTCTTGGT	CAATGAATTA	3900
GACATAGCAA	GTCTTCGCCA	GAAGTTTGAG	TCTACTGAAT	ATTTTATTCC	TAAGTCTGAA	3960
AAATCTTCC	TTGGTAAAGA	TAGAAATAAT	GTTGAATTGT	GGTTTGAAGA	AGTATGAAGT	4020
GGACCAAGAT	TATTAAAAAA	ATAGAAGAAC	AAATCGAGGC	AGGGATTTAT	CCCGGAGCCT	4080
CTTTTGCGTA	TTTTAAGGAC	AATCAATGGA	CAGAGTTCTA	TTTAGGCCAG	AGTGACCCAG	4140
AGCATGGCTT	GCAGACTGAG	GCAGGACTAG	TTTATGACCT	AGCTAGTGTC	AGCAAGGTTG	4200
TTGGGGTTGG	CACAGTTTGT	ACCTTCTTGT	GGGAAATAGG	TCAATTAGAT	ATTGATAGAC	4260
TGGTAATAGA	TTTTTTACCT	GAGAGTGATT	ATCCAGACAT	CACTATTGCG	CAGCTCTTGA	4320
CTCATGCAAC	AGACCTTGAT	CCTTTTATTC	CTAATCGTGA	TCTTTTAACA	GCCCCGAAT	4380
TAAAGGAAGC	GATGTTTCAT	CTCAACAGAC	GAAAGTCAGC	AGCCTTTCTT	TATTCGGATG	4440
TCCATTTTTT	GCTGTTGGGC	TTTATTTTGG	AAAGAATTTT	TAATCAAGAT	TTGGATGTGA	4500
TTTTAAAGGA	TCAAGTCTGG	AAACCTTGGG	GAATGACGGA	AACTAAGTTT	GGGCCAGTTG	4560
AGCTTGCTGT	TCCAACAGTT	AGAGGTGTAG	AGGCAGGCAT	AGTGCAATGAT	CCCAAGGCTC	4620
GTCTCCTGGG	TAGACATGCT	GGGAGTGCTG	GTTTATTTTC	GAATAAAAG	GATTTACAAA	4680
TCTTTTATGA	AACTATTTTA	GCAGATGATT	TTGCAAGAGA	CTTAAATCAA	AATTTTCTC	4740
CTTTGGATGA	CAAGGAACGT	TCTTTAGCAT	GGAATTTGGA	AGGAGATTGG	CTAGACCATA	4800
CGGGCTATAC	AGGTACCTTT	ATCATGTGGA	ATCGTCAGAA	GCAAGAAGCC	ACTATTTTCC	4860
TATCGAATCG	TACCTATGAA	AAGGACGAGA	GAGCTCAATG	GATATTAGAC	CGCAATCAAG	4920
TGATGAACCT	GATTCGCAAA	GAAGAGTAAG	GAGAGACATG	TCAAATAGTT	TAAAAGGGAC	4980
TTTACTAACA	GTTGTGGCTG	GTATTGCTTG	GGGGTTGTCA	GGAACGAGTG	GCCAATACCT	5040

1035

AATGGCACAC GGAATTTCCG CTCTGGTCTT GACTAACTTG CGTCTTTTAA TCGCTGGTGG	5100
AATTCTCATG CTCTGGCTT ATGCTACTGC AAAGGATAAA ATACTGGTCT TTTTAAAGGA	5160
TAGAAAGAGT TTGCTGTCTC TTCTTATTTT TGCTCTGATT GGTCTTTTTC TCAACCAATT	5220
CGCCTATCTG TCTGCTATTC AGGAGACCAA TCGGGGAACA GCGACGGTGC TTCAGTATGT	5280
TTGTCCTGTC GGAATTTTAA TTTATAGCTG TATCAAGGAT AGGGTGGCAC CGACACTGGG	5340
AGAGATAGTT TCCATCATAT TCGCCATCGG AGGAACCTTC CTGATCGCAA CACATGGGCA	5400
GTTGGACCAG TTATCCATGA CACCTGCTGG TCTGTTCTGG GGTCTCTTTT CTGCCCTGAC	5460
TTATGCTCTG TATATCATTT TACCCATAGC CTTGATTAAA AAGTGGGGGA GCAGCTTGGT	5520
CATTGGTGTG GGAATGGTCA TAGCAGGTTT GGTGCGCCCT CTTTTTACAG GGGTCTTACA	5580
GGCCGATATC CCGACTAGTC TTGATTTTCT CTTGCGTTT GCAGGCATTA TCCTTATCGG	5640
GACTGTCTTT GCCTATACAG CTTTCCTTAA AGGAGCCAGT CTGATAGGAC CGGTCAAGTC	5700
AAGCTTGTTG GCTTCAATG AGCCAATATC GGCGATTTTC TTTGCCCTTCT TAATAATGAA	5760
TGAACAATTT TATCCCATG ATTTTCTTGG TATGGCAATG ATATTGTTT CTGTAACTTT	5820
GATTTCTTTG AAAGATTTAT TCTTAGAAAA ATAAAAAGA CTCTTTGTCC GTGACAGAGA	5880
GTTTTTGCGT GGAATCTAA TTATTTTCAA GATAAAATTC AAAGCGTTCC CCTACATATT	5940
GACTTTTTAC GTATTCAAAA GCAGTACCAT CTTCTAGGTA GGAAACCTGG GTCAATCCAA	6000
GAATAGCATG TCCTTTTTC AACTCCAAAT AGTGGGCAAT CTTTTCTTTA GCAAGGCGAG	6060
CATAGATGGT CTGTTGAGAT TTGCCGATAC GATAGCCATG TTTTTCGAAG GTTTGGAAGA	6120
AATGACTGGT GATTTCTTCT TTTTAAAGT CCTTAATGAA TTTTTCAGGA ATAGAAGCAA	6180
CTTCATAAAC TAGGGGAACT TGGTCGGCAT AGCGGACCCG CTCCATTCGG ATAATATTGT	6240
CCGTTGGAAA AATTCCTAGC TTGGCAACTT CTTGCTCATT GGAATGGTT TTTTGTAGG	6300
AAATGAGCTG GCTAGAGGGA ACTTTACCTT GGGATTTGAC AATTCAGTA AAATGGTTG	6360
TCCCTCGCAT CTTTCTTGT ACTCGAGTAC TGGAAACAAA GGTGCCGCTT CCTACACGGC	6420
GCTCTAAGAC GCCTTCTTCG ACTAATAGAG ATACGGCTTG GCGGAGGGTC ATGCGACTGA	6480
CCGCAAACTG CTCAGCTAAA TCTCTTTCAC TGGGAAGCCT CTCACCAATA GCCCAACGGT	6540
ACTCGTCAAT ATCCTTTTTT ATCTGATCAT GGAITTTTAT ATAAGCAGGT AGCATATTTT	6600
TCACTTCATT TCTATCTTTT CTCTATTGTA CCCCAATAAA CTAGAAAAAG TCAAACTTCG	6660
CCTTGTTTAG TTGGTAATTC GCCCTTATTT GTGATAGAAT ATTGAGAAAA GATATTTCTT	6720
TTGAGAAAGG AAAAAGATGA GCAACATTC AACTGATTTG CAAGATGTAG AAAAAATCAT	6780

1036

CGTATTGGAC TATGGTAGCC AGTACAACCA GCTGATTTCA CGCCGTATCC GTGAGATTGG	6840
TGTTTTTTCA GAACTAAAAA GCCATAAAAT TTCAGCTGCT GAAGTTCGTG AAGTCAATCC	6900
TGTAGGAATT ATTCTATCAG GTGGTCCAAA TTCTGTATAT GAAGATGGTT CATTTGATAT	6960
TGACCCAGAA ATCTTCGAAC TCGGAATTCC AATTTGGGA ATCTGTTATG GTATGCAGTT	7020
ATTGACCCAT AAACCTGGAG GAAAAGTTGT TCCTGCAGGT GATGCTGGAA ATCGTGAATA	7080
CGGTCAATCA ACCCTAACTC ACACACCATC AGCGCTTTTT GAATCAACAC CTGATGAACA	7140
GACTGTTTTG ATGAGCCATG GTGATGCGGT TACTGAGATT CCTGCTGACT TTGTTCTGAC	7200
AGGTACATCA GCTGACTGCC CATACGCAGC CATCGAAAAC CCAGATAAAC ACATTACGG	7260
TATCCAATTC CACCCAGAAG TTCGTCATTC TGTATACGGA AATGATATCC TTCGTAACCT	7320
TGCCCTTAAC ATTTGTAAGG CTAAAGGTGA CTGGTCAATG GATAATTTCA TTGACATGCA	7380
GATCAAAAAA ATTCGTGAAA CCGTCGGTGA TAAACGTGTC CTTCTTGGTC TATCAGGTGG	7440
TGTTGACTCA TCTGTCGTTG GGGTCTTCT CCAAAAAGCG ATTGGCGATC AATTGATCTG	7500
TATCTTCGTA GACCACGGTC TTCTTCGTAA AGGCGAAGCT GATCAAGTTA TGGACATGCT	7560
CGGTGGTAAG TTTGGTTTGA ATATCGTCAA AGCAGACGCT GCTAAACGTT TCCTTGACAA	7620
ACTTGCTGGC GTTTCTGACC CTGAACAAAA ACGTAAAATC ATCGGTAACG AGTTTGTCTA	7680
TGTATTGAT GACGAAGCAA GCAAGCTCAA AGATGTGAAA TTCCTTGCTC AAGGTACTTT	7740
ATATACAGAT GTTATCGAGT CTGGTACGGA TACAGCTCAA ACTATCAAGT CACACCACAA	7800
CGTGGGCGTC TTCCAGAAGA TATGCAGTTT GAATTGATTG AACCACCTCAA TACTCTTTAC	7860
AAGGATGAAG TTCGTGCTCT TGGTACAGAG CTTGGTATGC CAGACCATAT CGTATGGCGC	7920
CAACCATTCC CAGGACCAGG ACTTGCTATC CGTGTCATGG GTGAAATCAC TGAAGAGAAA	7980
CTTGAAACCG TTCGTGAATC AGACGCTATT CTTGCTGAAG AAATCGCTAA AGCTGGACTT	8040
GACCGCGATA TTTGGCAATA CTTCACTGTT AACACAGGCG TTCGTTCACT CGGTGTTATG	8100
GGTGACGGTC GTACGTATGA CTACACGATT GCAATCCGTG CTATCACTTC TATCGATGGT	8160
ATGACTGCTG ATTTTGCCAA AATTCCATGG GAAGTACTTC AAAAAATCTC AGTACGTATC	8220
GTAAATGAAG TGGATCATGT TAACCGTATC GTCTACGATA TTACAAGTAA ACCACCTGCA	8280
ACAGTTGAGT GGGAATAATC GCAAAAAAAT TAAAAGCTTT GTAAAATCAA CGGTTACAGA	8340
GGATTAAAAA CTGTAACCTG GATTAAAAAG GGAACATTG CTAAAAAGAA TAAATTGAAT	8400
AATAGTTCCA AGTGGTTTAC ATTTGGACAA AAAATTAGAC CGTAGTTTTT AAGCTGCGGT	8460
CTTTTGATAT ATATAATGAG AATTAATGGC TCTTTGTCAA CTGTAGTGGG TTGAAGTCAG	8520
CTAAGCTCGA GAAAGGACAA ATTTGTCCT TTCTTTTTTG ATATTCAGAG CGATAAAAAAT	8580

1037

CCGTTTTTTG AAGTTTTCAA AGTTCCGAAA ACCAAAGGCA TTCCGCTTGA TAAGTTTGAT	8640
GAGATTATTG GTCGCTTCCA ATTTGGCGTT AGAATAGTGT AGTTGAAGGG CGTTGACGAT	8700
TTTCTCTTTG TCCTTTAGAA AGGTTTTAAA GACAGTCTGA AAAAGAGGAT GAACCTGCTT	8760
TAGATTGTCC TCAATGAGTC CGAAAAATTT CTCGGTTCC TTATTCTGAA AGTGAAACAG	8820
CAAGAGTTGA TAGAGCTGAT AGTGATGTTT CAAGTCTTGT GAATAGCTCA AAAGCTTGTT	8880
TAAAATCTCT TTATTGGTTA AATGCATACG AAAAGTAGGG CGATAAAAAT GTTTATCGCT	8940
GAGTTTACGA CTATCCTGTT GTATGAGCTT CCAGTAGCGC TTGATAGCCT TGTATTTCATG	9000
AGACTTTTGA TCCAATTGAT TCATGATTG AACACGCACA CGACTCGG	9048

(2) INFORMATION FOR SEQ ID NO: 160:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 10399 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

GTACCTTTAT TGATGAATGG ACTGTTTAAA TCAGTAGCAC GCCAACCAGA TATGCTTTCT	60
GAGTTTCGTA GTTTGATGTT TTTAGGTGTT GCCTTTATTG AAGGAACTTT CTTTGTAAC	120
CTTGCTTTCT CATTTATTAT CAAATAAATA CATGGAACGA GAAGAAAAGG GAGGATTTTA	180
GATGGAAGAA AGTATTAATC CAATCATCTC TATTGGTCCT GTTATCTTCA ATCTGACTAT	240
GTTAGCCATG ACTTTGTTGA TTGTGGGAGT TATTTTTGTC TTTATTTATT GGGCAAGCCG	300
CAATATGACC TTGAAACCCA AAGGAAAGCA AAATGTACTT GAGTATGTCT ATGACTTTGT	360
TATTGGATTT ACAGAACCTA ACATTGGTTC GCGCTACATG AAAGATTACT CACTCTTTTT	420
CCTTTGTTTA TTCCTTTTCA TGGTGATTGC CAATAACCTT GGCTTAATGA CAAAGCTTCA	480
AACGATCGAT GGGACTAACT GGTGGAGTTC GCCAACCGCT AATTTACAGT ATGACTTAAC	540
CTTATCTTTT CTTGTCATTT TGTGACACA TATAGAAAGC GTTCGTCGTC GTGGATTTAA	600
AAAAAGTATA AAATCTTTTA TGAGTCCTGT TTTTGTCATA CCGATGAATA TCTTGAAGA	660
ATTTACAAAC TTCTTATCTT TGGCTTTGCG GATTTTTGGG AATATCTTTG CAGGAGAGGT	720
CATGACGAGT TTGTTACTTC TTCTTTCCCA CCAAGCTATT TATTGGTATC CAGTAGCCTT	780
TGGAGCTAAT TTGGCTTGGA CTGCATTTTC TGTCTTTATT TCCTGCATCC AAGCTTATGT	840
TTTTACTCTT TTGACATCTG TGTATTTAGG GAATAAGATT AATATTGAAG AGGAATAGAA	900

1038

AGGAGTAACT GATGCACGTA ACAGTAGGTG AATTAATTGG TAATTTTATT TTAATCACTG	960
GCTCTTTTAT TCTTTTGCTA GTCTTGATTA AAAAATTTGC ATGGTCTAAT ATTACAGGCA	1020
TTTTCGAAGA AAGAGCTGAA AAAATTGCTT CAGATATTGA CAGAGCTGAA GAAGCCCGTC	1080
AAAAAGCAGA AGTATTGGCT CAAAACGCG AAGATGAATT GGCTGGTAGC CGTAAAGAAG	1140
CTAAGACAAT CATTGAAAAT GCAAAGGAAA CAGCTGAGCA AAGTAAGGCT AATATCTTAG	1200
CAGATGCTAA ACTAGAAGCA GGACACTTAA AAGAAAAAGC CAATCAAGAA ATTGCTCAAA	1260
ATAAAGTAGA AGCTTTACAG AGTGTTAAGG GTGAGGTCGC AGATTTGACC ATCAGCTTAG	1320
CTGGTAAAAT CATCTCACAA AACCTTGACA GTCATGCCCA TAAAGCACTC ATTGATCAGT	1380
ATATCGATCA GCTAGGAGAA GCTTAATGGA CAAGAAAACA GTAAAGGTAA TTGAAAAATA	1440
CAGCATGCCT TTTGTCCAAT TGGTACTTGA AAAAGGAGAA GAAGACCGTA TCTTTTCAGA	1500
CCTGACTCAA ATCAAGCAAG TTGTTGAAAA AACAGGTCTG CTTCTTTTTT TAAAACAAGT	1560
GGCAGTAGAC GAGTCGGATA AGGAAAAAAC AATTGCTTTT TTCCAAGATT CTGTGTCGCC	1620
TTTATTACAA AACTTTATCC AGGTTCTGGC CTACAATCAC AGAGCAAATC TTTTTTATGA	1680
TGTGCTTGTA GATTGCTTGA ACCGACTTGA AAAAGAAACA AATCGATTGG AAGTGACGAT	1740
TACGTCTGCT CATCCTCTAA CTGATGAACA GAAGACTCGT TTGCTCCCTT TGATTGAGAA	1800
AAAAATGTCT CTGAAAGTAA GGAGTGTAAA AGAACAAATC GATGAAAGTC TCATTGGTGG	1860
TTTTGTCAAT TTTGCCAATC ACAAGACAAT TGATGTGAGT ATTAACAAC AACTTAAAGT	1920
TGTTAAAGAA AATTGAAAT AGAAAGTGGT GTTCTTTTGG CAATTAACGC ACAAGAAATC	1980
AGCGCTTTAA TTAAGCAACA AATTGAAAAT TTCAAACCCA ATTTTGATGT GACTGAAACA	2040
GGTGTTGTAA CCTATATCGG GGACGGTATC GCGCGTGCTC ACGGCCTTGA AAATGTCATG	2100
AGTGGAGAGT TGTGAATTT TGAACACGGC TCTTATGGTA TGGCTCAAAA CTTGGAGTCA	2160
ACAGACGTTG GTATTATCAT CCTAGGTGAC TTTACAGATA TCCGTGAAGG CGATACAATC	2220
CGCCGTACAG GGAAATCAT GGAAGTCCCT GTAGGTGAAA GTCTGATTGG TCGTGTGTG	2280
GATCCGCTTG GTCGTCCAGT TGACGGTCTT GGAGAAATCC AACTGATAA AACTCGTCCA	2340
GTAGAAGCAC CAGCTCCTGG TGTATGCAA CGTAAGTCTG TTTCAGAACC ATTGCAAAC	2400
GGTTTGAAAG CTATTGACGC CCTGTACCG ATTGGTCGTG GTCAACGTGA GTTGATTATC	2460
GGTGACCGTC AGACAGGGAA AACAACCATT GCGATTGATA CAATCTTGAA CCAAAAAGAT	2520
CAAGATATGA TCTGTATCTA CGTCGCGATT GGACAAAAAG AATCAACAGT TCGTACGCAA	2580
GTAGAAACAC TTCGTCAGTA CGGTGCCTTG GACTACACAA TCGTTGTGAC AGCCTCTGCT	2640
TCACAACCAT CTCCATTGCT CTTCTAGCT CTTATGCTG GGGTTGCTAT GGCGGAAGAA	2700

1039

TTTATGTATC AAGGTAAGCA TGTTTTGATT GTATACGATG ATCTATCAAA ACAAGCGGTA	2760
GCTTATCGTG AACTGTCGCT CTTGCTTCGT CGTCCTCCAG GTCGTGAAGC CTTCCCAGGG	2820
GATGTTTTCT ATCTCCACAG CCGTTTGCTT GAGCGCTCAG CTAAAGTTTC TGATGAACTT	2880
GGTGGTGGAT CAATTACAGC CCTACCATTT ATCGAGACAC AAGCAGGAGA TATCTCAGCC	2940
TATATCGCAA CCAACGTGAT TTCTATCACT GATGGACAAA TCTTCCTTGG CGATGGCCTC	3000
TTCAATGCAG GTATTGCTCC AGCCATCGAT GCGGGTTCAT CTGTATCTCG TGTAGGTGGT	3060
TCTGCACAAA TCAAAGCCAT GAAGAAGGTT GCTGGTACAC TTCGTATCGA CCTTGCTTCA	3120
TACCGTGAGT TGGAAAGCCTT TACTAAGTTT GGTTCGACT TGGACGCAGC AACACAGGCT	3180
AAGTTGAACC GTGGACGTCG TACCGTTGAG GTCTTGAAAC AACCTGTTCA CAAACCATTA	3240
CCTGTTGAGA AACAAGTAAC CATTCTTTAT GCTTTGACAC ATGGTTTCTT GGATACTGTT	3300
CCAGTAGATG ATATTGTTCC TTTTCGAGGAA GAGTTCATG CCTTCTTTGA TGCTCAACAT	3360
CCAGAGATTT TGGAAACCAT TCGTGATACA AAAGACTTGC CAGAAGAAGC AGTCTTGAT	3420
GCTGCGATTA CAGAGTTTCT CAATCAATCT AGCTTCCAAT AAGAATAGAG GTGTCAGATG	3480
GCAGTATCTC TAAATGATAT TAAAACAAAA ATCGCCTCAA CAAAAATAC GAGTCAAATC	3540
ACTAATGCCA TGCAAATGGT ATCGGCTGCT AAGCTAGGTC GTTCTGAAGA AGCTGCTCGC	3600
AACTTCCAAG TTTACGCTCA GAAAGTGCCT AAACTTTGA CAGATATCCT TCATGGTAAT	3660
GGAGCTGGTG CTTCAACTAA TCCGATGTTG ATTAGCCGTT CTGTGAAGAA GACAGGCTAT	3720
ATCGTTATCA CTTCAGACCG CGGTTTGTTT GGAGGTTATA ATTCCTCTAT TTTGAAAGCT	3780
GTTATGGAGT TGAAAGAAGA ATACCACCCA GACGGTAAAG GTTTTGAAAT GATCTGTATC	3840
GGTGGGATGG GAGCTGATTT CTTTAAGGCT CGCGGTATTC AACCACTTTA TGAATTACGT	3900
GGCTTGTCAG ACCAACCTAG CTTTGATCAA GTTCGTAAGA TTATTTCAAA AACTGTTGAA	3960
ATGTACCAA ATGAACCTCT TGATGAGCTT TATGTTTGCT ACAACCACCA TGTCATACG	4020
CTAACCAGTC AAATGCGTGT GGAACAAATG CTTCCGATTG TTGACTTGGA TCCAAATGAA	4080
GCGGATGAAG AGTACAGCTT GACTTTTGAA TTGGAAACCA GCCGAGAAGA AATTCTGGAG	4140
CAGTTGTTGC CTCAGTTTGC AGAAAGTATG ATTTACGGTG CCATTATCGA TGCCAAGACA	4200
GCTGAGAATG CTGCGGGCAT GACAGCCATG CAAACAGCGA CAGATAATGC TAAGAAAGTC	4260
ATCAATGATT TGACAATTCA GTATAACCGT GCCAGACAGG CGGCGATTAC ACAAGAAATT	4320
ACAGAAATCG TAGCAGGTGC TAGTGCCTTA GAATAGGCTC TAGTCCAGCT CGTATGAAAA	4380
TGAACCTTAGG ACCTAGTTGA GCTAGGAACC GACAGTATCT TATATAGAAT AGGAGAAGGA	4440

1040

GATGAGTTCA CGTAAAATTG CTCAGGTTAT CGGTCCCCTT GTAGACGTTT TGTTCGACGC	4500
AGGGGAAAAA CTTCTGAGA TTAACAATGC ACTTGTCGTC TACAAAAATG ACGAAAGAAA	4560
AACAAAAATC GTCCTTGAAG TAGCCTTGA GTTAGGAGAT GGTATGGTTC GTACTATCGC	4620
CATGGAATCA ACAGATGGGT TGA CTCTGACCA GTAGGTAAAG AAACCTTGGG ACGTGTCTTC AACGTTTGG GAGATACCAT	4680
CTCTGACCA GTAGGTAAAG AAACCTTGGG ACGTGTCTTC AACGTTTGG GAGATACCAT	4740
TGACTTGGAA GCTCCTTTTA CAGAAGACGC AGAGCGTCAG CCAATTCATA AAAAAGCTCC	4800
AACTTTTGAT GAGTTGTCTA CCTCTTCTGA AATCCTTGAA ACAGGGATCA AGGTTATTGA	4860
CCTTCTTGCC CTTTACCTTA AAGGTGGTAA AGTTGGACTT TTCGGTGGTG CCGGAGTTGG	4920
TAAAACTGTC TTAATCCAAG AATTGATTCA CAACATTGCC CAAGAGCACG GTGGTATTTT	4980
AGTATTTGCT GGTGTTGGGG AACGTACTCG TGAGGGGAAT GACCTTTACT GGGAAATGAA	5040
AGAATCAGGC GTTATCGAGA AAACAGCCAT GGTCTTTGGT CAGATGAATG AGCCACCAGG	5100
AGCACGTATG CGTGTGCCCC TTA CTCTGATG GGTGTTGGT CAGATGAATG AGCCACCAGG	5160
AGGCAAGAC GTGCTTCTCT TTATCGATAA TATCTTCCGT TTCCTCAGG CTGGTTCAGA	5220
AGTATCTGCC CTTTGGGGTC GTATGCCATC AGCCGTGGT TACCAACCAA CACTTGCTAC	5280
GGAAATGGGT CAATTGCAAG AACGTATCAC ATCAACCAAG AAGGGTTCTG TAACCTCTAT	5340
CCAGGCTATC TATGTGCCAG CGGATGACTA TACTGACCCA GCGCCAGCAA CAGCCTTCGC	5400
TCACTTGGAT TCAACAACAA ACTTGGAACG TAAGTTGGTA CAATTGGGTA TCTACCCAGC	5460
CGTTGACCCA CTTGCTTCAA GCTCACGTGC CTTGGCACCT GAAATCGTTG GAGAAGAGCA	5520
CTATGCAGTT GCTGCTGAAG TAAAACGTGT CCTTCAACGT TACCATGAAT TGCAAGATAT	5580
CATTGCTATC CTTGGTATGG ATGAGCTTTC TGATGAAGAA AAGACCTGG TTGCTCGCGC	5640
CCGTCGTATC CAGTTCTTCT TGTACAAAA CTTCAACGTT GCGGAACAAT TTAAGGAAA TCCTTGATGG	5700
GCCAGGTTCT TATGTTCCAG TTGCTGAAAC TGTACGTGGC TTTAAGGAAA TCCTTGATGG	5760
TAAATACGAC CACTTGCCAG AAGATGCCTT CCGTGGTGTA GGTTCATATG AAGATGTGAT	5820
TGCAAAAGCT GAAAAAATGG GATTTTAAGA GGTGATCTAT GGCTCAGTTA ACTGTCCAGA	5880
TCGTGACACC AGATGGTCTC GTCTATGATC ACCATGCCAG CTATGTATCG GTTCGAACTC	5940
TGGATGGTGA GATGGGGATC TTGCCACGAC ATGAAAATAT GATTGCGGTT TTAGCAGTTG	6000
ATGAAGTAAA GGTAACACGT ATCGATGATA AAGATCACGT GAACTGGATT GCAGTAAACG	6060
GTGGCGTTAT TGAAATTGCC AATGATATGA TCACAATCGT CGTGACTCT GCAGAACGTG	6120
CTCGTGATAT CGATATCAGT CGTGCAGAAC GTGCCAACT TCGTGCAGAA CGTGCAATTG	6180
AAGAAGCACA AGACAAACAT TTGATTGACC AAGAACGTCG TGCTAAGATT GCTTTGCAAC	6240



1041

GTGCTATTAA CCGTATTAAT GTCGGAAATA GACTATAAGA AAAAATGAAC TTGAAAATAC	6300
CAAGTTCATT TTTTATGGTG TTTTAAGGAG CAAAACGGAT GCAGACTGCT TCGGGAACAT	6360
GGAAGTCGTT GGAGAGTTCT GCTAGACGAC CATTGTCACA ATTACGTTTA AAGACAGTTG	6420
CATTGTCAGA GTCTTGATGG ACAACAATGA GAAATTTTGT GTCGGGTGTC AAATCAAAAT	6480
CACGTGGAGT CTGACCATGC GTTGGAACGA TTTCTAATAA CTCTAAGCTA CCGTCCGCAA	6540
GGATGGTATA TACTGCGATA GAATCATGGC CACGGTTAGA AGCGTAGAGG TATTTACCGT	6600
CTTTAGAGAG ATGAATAGCA GCGGTTCCAT TAAAGCCTTC GTAAGCTTCC GGTAAAGTTG	6660
AAATGACCTG CATACTTCA AATTCGCCAA CGCCATCGTA GATTAAAACT TCGATAGTAC	6720
TATTGAGTTC ACAAATGAGA TAAGCGATTT TATAGTGGTT ATGGAAAATG ATATGCCGTG	6780
AGCCTGCTCC TGGCTTGCTG TGATAGGTAT AGAGCTTAGA TAATTTTCCT TCTTGATCGA	6840
GGTCATAGGT GATGACTTGG TCAGTTCCCA AGTCGCAGGT CACTAGATAG TGGTCAGGTG	6900
TTAAATCTGT ATAGTGAACA TGGGGGAAG CTTGATTTTC ATGTGGACCT TGGCCACTGT	6960
GTTGATCCAT ATCACTAAGT AGAAGACTAC CATCTTCCTG GCGTTTATAA ACAAGGACTT	7020
GTCCCTTG TG ATAGTTAGCT GCGTAAACCA AATCAGCCTT TTCATCGACA GCAACATAAC	7080
AGTGGGGAGC TCCTTCTTCA ACAACATGAT TTAACACAGT CCCGTCAGTT TGATAGGCTG	7140
CAATTCCTCC CTTATCGTCT TGGCTACCAA CAGTGATATA ATGTTGGTGC TGGTCAAAGC	7200
CAAGGTAGGT TGGACTTGGC TCAGCTGCAA AAAGTTCTAG ATTTGAAAGC TGACCAGTTT	7260
CTGTATCAAA GTCTGCCTTG TAAATCCCTT GAGAAGTACG ACGTGATATA GTTCCAAAAT	7320
AAACAGTTTC TTTCACTACT ATACCTCTGT GTAAAGATAA GACTATTATA TCACAAAAC	7380
AAGTAAATTA AAGATATCCA ATTAGATGTA AGCACTTTAA AAAAGAGTTA TTTGTTTCA	7440
AAAATGGTAT AATGAGAGAA CAATAGAAAG GAAGTATTTA TGGAGCAAAA AGAGAAACAT	7500
TTTAGCCTAT CTTGGTTTTT CAAGTGGTTT TTAGATAACA AGGCAATTAC GGTATTTTCA	7560
GTAACCTTAT TATTGGGACT GAATCTTTTT ATTTAAGTA AGATTAGTTT TCTATTTTCA	7620
CCTGTTTTAG ACTTTTTAGC AGTTGTGATG TTGCCAGTCA TTTGCTCTGG TTTGTTATAT	7680
TATTTGTTGA ATCCTATTGT TGATTGGATG GAGAAGCATA AGGTAAATCG TGTATAGCT	7740
ATCACTATTG TCTTTGTTAT CATCGCTCTC TTTATCATTT GGGGCTTGGC AGTCGCCATT	7800
CCAAATCTGC AACGTCAGGT TTTGACCTTT GCAAGAAACG TTCCTGTTTA CTAGAAGAT	7860
ATAGATAGGA TTGTTAATGG ATTGGTAGCC CAGCACCTGC CAGATGATTT CAGACCTCAA	7920
TTAGAGCAAG TTTTGACCAA TTTTCTAGC CAGGCTACAG TTTTGGCAAG TAAGGTTTCA	7980

1042

TCTCAGGCAG TCAACTGGGT GAGTGCCTTT ATTAGCGGGG CTTCTCAAGT GATTGTTGCC	8040
TTGATTATCG TTCCTTTCAT GCTCTTTTAT CTCTTGCGTG ATGGGAAAGG CTTGCGTAAC	8100
TATTTGACCC AATTCATTCC AAGAAAATTG AAGGAACCTG TTGGACAAGT TTTATCAGAT	8160
GTGAATCAAC AGTTGTCCAA CTATGTTTGA GGGCAAGTGA CAGTGGCTAT TATTGTAGCA	8220
GTAATGTTTA TCATCTTCTT CAAGATTATT GGTCTACGCT ATGCGGTTAC GCTGGGGGTT	8280
ACTGCTGGTA TTTTAAATCT GGTCCCTTAT CTTGGTAGCT TTCTAGCCAT GCTTCTCTGT	8340
CTAGTATTGG GTTTGATTGC TGGTCCAGTC ATGCTTTTGA AAGTAGTGAT TGTCTTTATC	8400
GTAGAACAAA CTATTGAAGG CCGTTTGTG TCTCCATTGA TTTTGGAAG TCAATTAAAC	8460
ATCCACCCTA TTAATGTTCT CTTTGTGTTG TTAACCTCAG GATCTATGTT TGGTATCTGG	8520
GGAGTTTAC TTGGTATTCC GGTTTATGCC TCTGCTAAGG TTGTCATTTT AGCCATTTTC	8580
GAATGGTATA AGGTAGTCAG TGGTCTATAT GAATTAGAGG GTGAGGAAGT CAAGAGTGAA	8640
CAATAGTCAA CAGATGTTAC AGGCTTTGGA GGAGCAAGAT TTAATAAGG CTGAGCATT	8700
TTTCGCCAAA GCTTTAGAAA ATGATTCAAG TGATCTTCTG TATGAATTGG CAACTTATCT	8760
TGAAGGGATT GGTTCCTATC CTCAGGCCAA GGAAATTTAC CTGAAAATTG TAGAGGATTT	8820
TCCAGAGGTT CATCTTAATC TAGCTGCAAT TGCTAGCGAG GATGGTCAAA TAGAAGAAGC	8880
CTTTACCTAT CTTGAGGAAA TCCAAGCTGA CAGTGACTGG TATGTCTCGT CTTTGGCTCT	8940
GAAGGCAGAC CTTTACCAGC TGAAGGTTT GACAGATGTG GCACGTGAGA AATTATTGGA	9000
GGCCTTGACC TACTCAGAGG ATTCTCTCTT GATATTGGGT TTGGCAGAGT TGGATAGTGA	9060
GTTGGAAAAT TACCAAGCGG CTATTCAAGC CTATGCCAG TTAGATAATC GCTCGATTTA	9120
TGAGCAAACG GGCATTTCCA CCTATCAACG AATTGGCTTT GCCTATGCTC AGTTAGGGAA	9180
ATTTGAAACG GCTACTGAGT TTTTAGAAAA AGCCCTGGAG TTAGAATACG ATGACTTAAC	9240
AGCTTTTGAG TTGGCCAGTC TTTATTTTGA TCAAGAAGAA TATCAAAAAG CCACCCTCTA	9300
CTTTAAGCAG CTTGATACCA TTTCTCCTGA CTTTGAAGGC TATGAGTATG GGTACAGTCA	9360
GGCTTTACAT AAGGAACATC AAGTTCAAGA AGCCCTGCGT ATCGCTAAGC AAGGATTAGA	9420
GAAAAATCCC TTTGAAACTC GCCTCTTGCT AGCTGCTTCA CAATTTTCTT ATGAATTGCA	9480
TGATGCTAGT GGTGCAGAAA ATTATCTCCT TACTGCAAAA GAAGACGCTG AGGATACAGA	9540
AGAAATCTTG CTTGTTTAG CCACTATTTA TCTGGAGCAG GAGCGTTATG AGGATATTCT	9600
AGAATTGCAG AGTGAGGAGC CAGAAAATCT TTTGACCAAG TGGATGATTG CTCGTTCTTA	9660
TCAAGAAATG GACGATTTGG ATACTGCTTA TGAGTATTAT CAAGAGTTGA CAGGAGATTT	9720
GAAGGACAAT CCAGAATTTT TGAACACTA TATCTATCTC TTGCGTGAAT TGGGACATTT	9780

1043

TGAAGAAGCA AAAGTCCATG CTCACACTTA CTTAAAACTG GTTCCAGATG ATGTGCAAAT	9840
GCAAGAAGCTG TTTGAGAGAT TGTAAGAATG TTAAACCCAA ATCATTGATA CCTCTCTCAA	9900
CTAGATGTAA CTTACAAAAC CCCTGACCTC ATGAGCCACT TTCTTCCTCC TCATGAGGTC	9960
AGTTTTACTT TCTGCTGTT CAGTATCGTT TTCTCTCGCT AGATTTCCTC AAAAGGGCAG	10020
ACTCCTCCCT TGGTGCGTCA CACGATTTTT TCATCTCGAC TGTTCCTTAA TGCATCATT	10080
ACGACGCTTT TCTTCTAGGT GGTTCATAAG GAACAGGAAG ATTCAGGTTG ACTTTTCTAA	10140
TCCTAGAATA AAGTGCTGAA AACAATTCGG AATAGGCATA GAGACTAGAC AATTTGAGGA	10200
GCTGCTTGGC TCCTGTTTGA ACACATTTTC CCACCACGTG AAGAAAAAGA TGGCGGAAGC	10260
GTTTGATTGT TAAAGTTTGG AAGTCACCTC CAGCTAGATG TTTGAGAAAA AGATAGAGAT	10320
TGTAGGCGAT ACAGCTCATC ATCATACGAA TTCGTTTTTG ATTAAGGTTG AACTATCCGT	10380
TTTATCGCCA AAAAATCGG	10399

(2) INFORMATION FOR SEQ ID NO: 161:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9409 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

GATAAGATTA AGTTAGAAAA GAAAGAACTA GGACATATCT ACCAGATTCA GGTTTTTAAT	60
AGCTATGGGC AGGAAGAAAT CTATCGTGTG ATTTTGATGG AGACCAATAT TAGTTCGGTT	120
TCAACCAATA TCAAGTATGC TGCTGTCTTG ATTAATACCA GTCAGTTGGA ACAGGCTAGT	180
CAAAAGCATG AGCAATTGAT TGTGGTCGTG ATGGCTAGTT TCTGGATTTT GTCTTTACTT	240
GCCAGTCTCT ATCTAGCTAG GGTCAAGTGT AGGCCCTGC TTGAGAGTAT GCAGAAGCAA	300
CAGTCTTTTG TGGAAATGC CAGTCATGAG TTACGAACTC CACTCGCAGT TTTGCAAAAT	360
CGCTTAGAGA CCCTTTTTTCG TAAGCCAGAA GCTACCATTA TGGATGTGAG CGAAAGCATT	420
GCATCGAGTT TGGAAGAACT CCGAAATATG CGTTTTTTAA CGACAAGCTT GCTGAACTTA	480
GCTCGGAGAG ATGATGGGAT TAAGCCGGAG CTTGCAGAAG TTCCAAGTAG CTTTTTTAAT	540
ACAACTTTCA CAACTACGA GATGATTGCT TCGGAAAATA ATCGTGTCTT CCGTTTTGAA	600
AATCGTATCC ATCGAACAAT TGTCACAGAT CAGCTTCTTC TGAACAACAT GATGACCATT	660
CTTTTCGATA ATGCCGTCAA GTATACTGAG GAGGATGGTG AAATTGATTT TCTTATCTCG	720

1044

GCGACCGATC	GCAATCTTTA	TTTACTTGTT	TCTGATAATG	GAATCGGTAT	TTCGACAGAA	780
GATAAAAAGA	AAATTTTGA	CCGTTTTTAT	CGAGTAGACA	AGGCTAGAAC	CCGGCAAAAA	840
GGTGGTTTTG	GTTTAGGATT	ATCCCTAGCC	AAGCAAATTG	TAGATGCTCT	AAAAGGAACT	900
GTTACTGTCA	AAGATAATAA	ACCCAAGGGA	ACAATCTTTG	AAGTGAAGAT	TGCCATTTCAG	960
ACACCATCTA	AAAAGAAAAA	ATAAAAATAT	CGCTCCAATT	GGGGCGATAT	TTTGGATTTA	1020
TCTTCTACGT	TTTCGTTTGA	TAATAGACCG	TTGAACTTT	AAAACAAGTA	AGCTGAATCC	1080
GATTGCTGCG	GCAAAGGCAA	GAGCAGTTGA	TAATTTTAAT	GCTAAAAAGA	TAAAACTAAA	1140
GATAGCAATA	CAGATACAAA	AAACAGCGAT	ATTAATAAAA	AATAGGATTT	CCTTGAGATT	1200
GGCATCAGAT	TGCGCTTCAG	GTGTATAAGC	TTGGTAATGA	GGAAGCTGCT	GGTTTAATTC	1260
TTCTTGATAG	TCTACCTCAT	AGGATTGTAA	TTTTCTTACG	GGCATGATTC	TCTCCTTAAC	1320
AGTACATACC	TATTTTATCA	TTTTTTCGGC	AGAGAATTAT	TACAGAAAGG	TTACAAAAAG	1380
AATAAAGTCC	CTTTTCATTT	TCAAAGCATG	GCTGATTTTG	GAGAAATGTG	GTATAATTTT	1440
TCTTATGGAA	AAGATTGTCA	TTACAGCAAC	TGCTGAAAGT	ATTGAACAAG	TTGAACAACT	1500
ACTCGAAGCT	GGCGTAGACC	GTATCTATGT	CGGTGAGAAA	GATTTTGGTC	TTCGTCTGCC	1560
AACGACCTTT	AGTTATGACC	AATTACGTGA	AATCGCTAAG	TTGGTTCATG	ATGCTGGTAA	1620
GGAATTGATC	GTTGCGGTCA	ATGCTCTCAT	GCACCAAGAT	ATGATGGACC	GTATCAAGCC	1680
TTTCTTAAAC	TTCTTGGAAG	AAATCAAGAC	AGACTATATT	ACGATTGGGG	ATGCAGGCGT	1740
CTTTTACGTA	GTTAACCGCG	ATGGTTATTC	ATTTAAGACC	ATCTACGATG	CTTCAACCAT	1800
GGTAACTAGC	AGTCGTCAGA	TTAACTTCTG	GGGACAAAAG	GCTGGCGCAT	CTGAGGCTGT	1860
TTTGGCGCGT	GAAATTCCAT	CAGCTGAACT	TTTCAAAATG	CCAGAGATTT	TGGAAATTCC	1920
TGCTGAAGTT	TTGGTTTACG	GTGCTAGCGT	CATCCATCAT	TCTAAACGTC	CACTCTTGCA	1980
AAACTACTAT	AACCTTACAC	ATATCGATGA	TGAAAAGACG	CATAAACGTG	ACCTCTTCTT	2040
GGCTGAGCCA	AGTGATCCAG	AGAGCCACTA	TTCCATTTTT	GAAGATAATC	ATGGGACCCA	2100
TATCTTTGCC	AACAATGACC	TTGATTTGAT	GATCAAAATTA	ACAGAAATTGG	TGGAGCATGG	2160
CTTTACTCGC	TGGAAACTAG	AAGGGCTCTA	CACTCCTGGT	CAGAACTTTG	TTGAGATTGC	2220
AAAACTCTTT	ATCCAAGCGC	GTAGCTTGAT	TCAAGAGGGC	AACTTTAGTC	ATGCTCAAGC	2280
CTTCTTGCTG	GATGAAGAAG	TTCGTAAACT	TCACCCTAAA	AACCGTTTCC	TTGATACAGG	2340
ATTTTATGAC	TACGATCCTG	ACATGGTTAG	ATAAAATACA	TGATTTCGTTG	AGAGAAGGAA	2400
GATGCAAAAC	TTTCTTCTCT	CAATTTTTCG	TATTTCTTCA	CTATTTTACA	AAAATCAGCA	2460
GGCTAGAATG	CTCTATTCTGA	TGGGATTTTT	AAGAAAAGTA	GTGTTCTTGA	GTTTGAAAAAT	2520

1045

TATCCTATGT TTGCAGGTGC CAAATGGCCC TTTTITGGT ATAATTTTTT ATAATGAAAA	2580
CGATTGGTAA TCGCTATGTT GTGGTGGATT TAGAGGCAAC TAGCACAGGT AGTAAGGCTA	2640
AAATTATCCA AGTGGGAATT GTCGTGATTG AGGACGGAGA AATCGTCGAT CACTATACGA	2700
CGGATGTCAA TCCACATGAA CCCTTGGATG CTCATATCAA AGAACTGACA GGATTGACAG	2760
ACCAACGTCT GCGCAAGCA CCTGATTTTT CGCAAGTTGC CAGAAAAATA TTTGACTTGG	2820
TGGAGGATGG GATTTTTGTA GCCCATAATG TTCAGTTTGA TGCTAATCTC TTGGCGGAAA	2880
ATTTATTTTT TGAAGGCTAT GAGCTAAGAA ACCCTCGTGT TGATACGGTC GAATTGGCCC	2940
AGGTCTTTTT CCCTGAACTG GAAAAATATA GCTTGCCGAT TTTGTGTCTA GAATTAGGAA	3000
TTCTCTTAA ACACGCACAC ACAGCCCTTT CAGATGCCCA AGCTACAGCA GAATTACTTC	3060
TTTTTTTACG GAAAAAGATG ACCCAGCTTC CTAAGGTCT CTTGGAACGC TTGCTGGAAA	3120
TGGCTGACGC TCTCCTATAT GAGTCTACC TGGTTATTGA GGAACTTAT CGCAACCAAT	3180
CTATCCTGAG TTCTCCAGAC TTGGTCCAAG TTCAAGGTCT ATATTTTAAG AAAACGGAAG	3240
CTTCTCTGGA GCCACGAAAA CTATCTCAAG ACTTTTCTAA AAATATTTCT CTGTTGAACC	3300
TTGAAGTGAG GGAGGAACAA GAAAGTTTGT CTAAGAGGT TGGCTTGCTA TTGAAAGATG	3360
AACCTGTCTC TCTGATTCAA GCGCCGACAG GGATTGGGAA AACCTATGGC TATCTCTTAC	3420
CCGCTTTATC TCAATCCAAA GAGCGACAAA TTGTTCTTAG TGTCCGACA AAGATTCTTC	3480
AAAATCAAAT CATGGAAGAA GAAGGTAAAC GCCTCAAGGA AGTGTCCAT ACAGATATTC	3540
ATAGCTTAAA GGGACCACAA AATTATCTGA AGTTGGATGC CTTTTATCAT TCCTTGACAG	3600
AAAATGATGA AAATCGCTTA TTTAGACGCT TTAATATGCA AGTCTGGTC TGGCTTACTG	3660
AGACAGAGAC AGGAGATTTG GATGAAATCG GGCAACTCTA CCGTTACCAA CATTTTCTAG	3720
CAGACCTTCG TCATGATGGG AATTTATCAT CCCAGAGCTT ATTTGTGACG GAAGATTTTT	3780
GGAAACGTAG TCAAGAAAGG GCAGAGACTT GCAAGCTTTT AGTGACTAAT CATGCCTATC	3840
TCGTAACCAG ACTTGAAGAT AATCCTGAAT TTGTCAGTGA CCGTTTACTG ATTATTGATG	3900
AAGTCCAAAA GATTTTGTGA GCTCTAGAAA ATCTGCTTCA AGAGACCTAC GATATACAAT	3960
CTATTATCGA TTTAATTGAT AAGGCTTTAG TAGGAGAAGA AACAGGGTT CAACAACGGA	4020
TACTAGAAAG TATTCGCTTT GAGTGTCTCT ACTTGATAGA ACAATTCAG TCTGGCAAAT	4080
CTAGGAAAAA TATCTTAGAT TCTCTGGACA ATCTCCATCA GTATTTTCA GAATTGGAAG	4140
TAGAAGACTT TGATGAGCTG GTTCGCTATT TTACAGCTGA AGGTGATTAC TGGCTTGAAG	4200
TAACTGAAAC GAGTCAAAAG AAAATTCAGA TTTCTTCTAC AAAATCAGGC CGTACTCTTC	4260

1046

TGTCCTCTTT ACTTCCTGAG AGTTGCCAAG TCTTGGGAGT ATCGGCTACT CTTGAGATTA	4320
GTCAGAGGGT TTCTTTGGCA GACCTTTTAG GCTATCCTGA AGCTAAATTT GTCAAGATTG	4380
AATCTCGGGG AAAACAGGAA CAAGAAGTGG TCATGGTCAA AGATTTCCCT CTGGTAACAG	4440
AAACCTCCTT AGAAGTCTAT GCCAGAGAGG TAGCTGCTTT ACTAGTGGAA ATTCAAGCTT	4500
TCCAGCAACC GATTTTGGTT CTCTTTACCG CTAAAGACAT GCTTCTAGCA GTATCGGATT	4560
TACTTACAGT TAGCCACTTG GCCCAGTATA AAAATGGGGA TGTTTCATCAG CTAAAGAAAC	4620
GCTTTGAAAA AGGTGAACAA CAAATCTTGC TTGGTGCAGC AAGTTTCTGG GAGGGAGTTG	4680
ATTTTTCAG CCATCCTTCT GTGATTCAAG TTGTACCGAG GCTTCCTTTC CAAAATCCTC	4740
AAGAACCCTT GACGAAAAAG ATTAATCAAG AACTGAATCA AGAAGGGAAA AATGCCTTTT	4800
ATGATTATCA ATTGCCAATG GCCATTATTC GTTTAAACA GGGTTTGGGA AGAAGTATGA	4860
GACGTGAATA CCAACGTTCC TTAACCTCTA TTTTGGATAG GAGAATCGTC GGAAAACGAT	4920
ACGGCAAACA AATAGTAGCA TCTCTAGCAG AAGAAGCGAC TGTTAAAACC ATCTCTCGAT	4980
CCGAAGTTGA CGAGGCTATT GATAGATTTT TTAATGAGCT TTGATAAATA GTATTGTATG	5040
AAAGTATAAG GTTAGTATAT ATGAAACGTT CTCTCGACTC AAGAGTCGAT TACAGTTTGC	5100
TCTTGCCAGT ATTTTTTCTA CTGGTCATCG GTGTGGTGGC TATCTATATA GCCGTTAGTC	5160
ATGATTATCC CAATAATATT CTGCCCATTT TAGGGCAGCA GGTCGCCTGG ATTGCCTTGG	5220
GGCTTGTGAT TGGTTTTGTG GTCATGCTCT TTAATACAGA ATTTCTTTGG AAGGTGACCC	5280
CCTTTCTATA TATTTTAGGC TTGGGACTTA TGATCTTGCC GATTGTATTT TATAATCCAA	5340
GCTTAGTTGC ATCAACGGGT GCCAAAACT GGGTATCAAT AAATGGAATT ACCCTATTCC	5400
AACCGTCAGA ATTTATGAAG ATATCCTATA TCCTCATGTT GGCTCGTGTC ATTGTCCAAT	5460
TTACAAAGAA ACATAAGGAA TGGAGACGCA CGGTTCCGCT GGACTTTTGG TTAATTTTCT	5520
GGATGATTCT CTTTACCATT CCAGTCCTAG TTCTTTTAGC ACTTCAAAGT GACTTGGGGA	5580
CGGCTTTGGT TTTTGTAGCC ATTTTCTCAG GAATCGTTTT ATTATCAGGG GTTTCTTGGC	5640
AAATTATTAT CCCAGTATTT GTGACTGCTG TAACAGGAGT TGCTGGTTTC TTAGCTATCT	5700
TTATTAGCAA GGACGGACGA GCTTTTCTTC ACCAGATTGG AATGCCGACC TACCAAATTA	5760
ATCGGATTTT GGCTTGGCTC AATCCCTTTG AGTTTGCCCA AACAACGACT TACCAGCAGG	5820
CTCAAGGGCA GATTGCCATT GGGAGTGGTG GCTTATTGG TCAGGGATTT AATGCTTCGA	5880
ATCTGCTTAT CCCAGTTCGA GAGTCAGATA TGATTTTAC GGTATTGCA GAAGATTTTG	5940
GCTTTATTGG CTCTGTCCTG GTTATTGCCC TCTATCTCAT GTTGATTAC CGTATGTTGA	6000
AGATTACTCT TAAATCAAAT AACCAGTTCT ACACTTATAT TTCCACAGGT TTGATTATGA	6060

1047

TGTTGCTCTT CCACATCTTT GAGAATATCG GTGCTGTGAC TGGACTACTT CCTTTGACGG	6120
GGATTCCCTT GCCTTTCATT TCGCAAGGGG GATCAGCTAT TATCAGTAAT CTGATTGGTG	6180
TTGGTTTGCT TTTATCGATG AGTTACCAGA CTAATCTAGC TGAAGAAAAG AGCGGAAAAG	6240
TCCCATTCAA ACGGAAAAG GTTGTATTAA AACAAATTAA ATAAGGAGAA AATCATGGTA	6300
AAAGTAGCAG TTATATTAGC TCAGGGCTTT GAAGAAATTG AAGCCTTGAC AGTTGTAGAT	6360
GTCTTGCGTC GAGCCAATAT CACATGTGAT ATGGTTGGTT TTGAAGAGCA AGTAACGGGT	6420
TCGCATGCAA TCCAAGTAAG AGCAGATCAT GTCTTTGATG GAGATTTATC AGACTATGAT	6480
ATGATTGTTT TCCTGGAGG TATGCCTGGT TCTGCACATT TACGTGATAA TCAGACCTTG	6540
ATTCAAGAAT TGCAAAGCTT CGAGCAAGAA GGAAGAAAAC TAGCAGCCAT TTGTGCGGCA	6600
CCAATTGCCC TCAATCAAGC AGAGATATTG AAAAAATAAGC GATACACTTG TTATGACGGC	6660
GTTCAAGAGC AAATCCTTGA TGGTCACTAC GTCAAGGAAA CAGTAGTGGT AGATGGTCAG	6720
TTGACAACCA GTCGGGGTCC TTCAACAGCC CTGCTTTTG CCTACGAGTT GGTGGAGCAA	6780
CTAGGAGGGG ACGCAGAGAG TTTACGAACA GGAATGCTCT ATCGAGATGT CTTTGGTAAA	6840
AATCAGTAAA ACGGGAGTTA TTCTCTCGTT TTTTATGTGG AAAACTCAGG GAAATCATCG	6900
CTTTTTTCAT AAAAAATGC TATAATGAAG GGTATGAAAT ATCACGATTA CATCTGGGAT	6960
TTAGGTGGAA CTTTACTGGA TAATTATGAA ACTTCAACAG CTGCATTGTG TGAAACATTG	7020
GCÀCTGTATG GTATCACACA AGACCATGAC AGTGTCTATC AAGCTTTAAA GGTTCCTACT	7080
CCTTTTGCGA TTGAGACATT CGCTCCCAAT TTAGAGAATT TTTTAGAAAA GTACAAGGAA	7140
AATGAAGCCA GAGAGCTTGA ACACCCGATT TTATTTGAAG GAGTTTCTGA CCTATTGGAA	7200
GACATTTCAA ATCAAGGTGG CCGTCATTTT TTGGTCTCTC ATCGAAATGA TCAGGTTTTG	7260
GAAATTTTAG AAAAAACCTC TATAGCAGCT TATTTTACAG AAGTGGTGAC TTCTAGCTCA	7320
GGCTTTAAGA GAAAGCCAAA TCCCGAATCC ATGCTTTATT TAAGAGAAAA GTATCAGATT	7380
AGCTCTGGTC TTGTCATTGG TGATCGGCCG ATTGATATCG AAGCAGGTCA AGCTGCAGGA	7440
CTTGATACCC ACTTGTTTAC CAGTATCGTG AATTTAAGAC AAGTATTAGA CATATAAGAA	7500
AAAGGAATAA GATGACAGAA GAAATCAAAA ATCTGCAGGC ACAGGATTAT GATGCCAGTC	7560
AAATTCAAGT TTTAGAGGGC TTAGAGGCTG TTCGTATGCG TCCAGGGATG TACATTGGAT	7620
CAACCTCAAA AGAAGGTCTT CACCATCTAG TCTGGGAAAT TGTTGATAAC TCAATTGACG	7680
AGGCCTTGGC AGGATTTGCC AGCCATATTC AAGTTTTTAT TGAGCCAGAT GATTTCGATTA	7740
CTGTTGTGGA TGATGGGCGT GGTATCCCAG TCGATATTCA GGAAAAACA GGCCGTCCTG	7800

1048

CTGTTGAGAC CGTCTTTACA GTCCTTCACG CTGGAGGAAA GTTCGGCCGT GGTGGATACA	7860
AGGTTTCAGG TGGTCTTCAC GGGGTGGGGT CGTCAGTAGT TAATGCCCTT TCCACTCAAT	7920
TAGACGTTCA TGTTCACAAA AATGGTAAGA TTCATTACCA AGAATACCGT CGTGGTCATG	7980
TTGTGCGAGA TCTTGAAATA GTTGGAGATA CGGATAAAAC AGGAACAACCT GTTCACTTCA	8040
CACCGGACCC AAAAATCTTC ACTGAAACAA CAATCTTTGA TTTTGATAAA TTAAATAAAC	8100
GGATTCAAGA GTTGGCCTTT CTAAATCGCG GTCTTCAAAT TTCAATTACA GATAAGCGCC	8160
AAGGTTTGA ACAAACCAAG CATTATCATT ATGAAGGTGG GATTGCTAGT TACGTTGAAT	8220
ATATCAACGA GAACAAGGAT GTAATCTTTG ATACACCAAT CTATACAGAC GGTGAGATGG	8280
ATGATATCAC AGTTGAGGTA GCCATGCAGT ACACAACCTGG TTACCATGAA AATGTCATGA	8340
GTTTCGCCAA TAATATTCAT ACCCATGAAG GTGGAACACA TGAACAAGGT TTCCGTACAG	8400
CCTTGACACG TGTATCAAC GATTATGCTC GTAAAAATAA GTTACTGAAA GACAATGAAG	8460
ATAATTTAAC AGGGGAAGAT GTTCGCGAAG GCTTAACTGC AGTTATCTCA GTTAAACACC	8520
CAAATCCACA GTTTGAAGGA CAAACCAAGA CCAAATGGG AAATAGCGAA GTGGTCAAGA	8580
TTACCAATCG CCTCTTCAGT GAAGCTTTCT CCGATTTCTT CATGGAAAAT CCACAGATTG	8640
CCAAACGTAT CGTAGAAAAA GGAATTTTGG CTGCCAAGGC TCGTGTGGCT GCCAAGCGTG	8700
CGCGTGAAGT CACACGTAAG AAATCTGGTT TGGAAATTTT CAACCTTCCA GGGAAACTAG	8760
CAGACTGTTT TTCTAATAAC CCTGCTGAAA CAGAACTCTT CATCGTCGAA GGAGACTCAG	8820
CTGGTGGATC AGCCAAATCT GGTCGTAACC GTGAGTTTCA GGCTATCCTT CCAATTCGCG	8880
GTAAGATTTT GAACGTTGAA AAAGCAAGTA TGGATAAGAT TCTAGCCAAC GAAGAAATTC	8940
GTAGTCTTTT CACAGCCATG GGAACAGGAT TTGGCGCAGA ATTTGATGTT TCGAAAGCCC	9000
GTTACCAAAA ACTCGTTTTG ATGACCGATG CCGATGTGCA TGGAGCCAC ATTCGTACCC	9060
TTCTTTTAAC CTTGATTTAT CGTTATATGA AACCAATCCT AGAAGCTGGT TATGTTTATA	9120
TTGCCCAACC ACCAATCTAT GGTGTCAAGG TTGGAAGCGA GATTAAAGAA TATATCCAGC	9180
CGGGTGCAGA TCAAGAAATC AAATCCAAG AAGCTTTAGC CCGTTATAGT GAAGGTCGTA	9240
CCAAACCGAC TATTCAGCGT TATAAGGGGC TAGGTGAAAT GGACGATCAT CAGCTGTGGG	9300
AAACAACCAT GGATCCCGAA CATCGCTTGA TGGCTAGAGT TTCTGTAGAT GATGTGCAGA	9360
AGCAGATAAA ATCTTTGATA TGTTGATGGG GATCGAGTTG TCCTCGTCG	9409

## (2) INFORMATION FOR SEQ ID NO: 162:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6415 base pairs  
 (B) TYPE: nucleic acid



1049

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

CCTGGGAAAG TCTTGAAAAT TATGATAGAA TGGTGGAAGG AAAAATTCAG GAGAGTAGTA	60
GTGACTCAAA ATGTTGAAAG TCTTCTCGTA TCCATTGTAA TCAGTGCATA CAATGAAGAA	120
AAATATCTGC CTGGTCTAAT TGAAGACTTA AAAAATCAAA CCTATCCTAA AGAGGATATT	180
GAAATTCTAT TTATAAATGC TATGTCCACA GATGGGACCA CAGCTATCAT TCAGCAATTT	240
ATAAAGGAAG ATACAGAGTT TAACTCAATT AGATTGTATA ACAATCCTAA GAAAAATCAA	300
GCTAGTGGTT TTAACCTGGG AGTTAAACAT TCTGTAGGGG ACCTTATTTT AAAAATTGAT	360
GCTCATTCAA AAGTTACTGA GACTTTTGTA ATGAACAATG TGGCTATTAT TCAACAAGGT	420
GAATTTGTCT GTGGGGGGCC TAGACCGACG ATTGTGGAAG GAAAAGGAAA ATGGGCAGAG	480
ACCTTGCATC TTGTTGAGGA AAATATGTTT GGCAGTAGCA TTGCCAATTA TCGAAATAGT	540
TCTGAGGATA GATATGTTTC TTCTATTTT CATGGAATGT ATAAACGAGA GGTTTTCCAG	600
AAGTTTGGTT TAGTAAATGA GCAACTTGGC CGAACTGAAG ATAATGATAT TCATTATAGA	660
ATTCGAGAAT ATGGTTATAA AATCCGCTAT AGCCCAAGTA TTCTATCTTA TCAGTATATT	720
CGACCAACAT TCAAGAAAAT GCTGCATCAA AAGTATTCAA ATGGTTTGTG GATTGGCTTG	780
ACAAGTCATG TTCAGTTTAA GTGTTTATCA TTATTTCACT ATGTTTCTTG TTTATTTGTT	840
TTGAGTCTTG TGTTTAGTCT AGCATTGTTA CCGATCACAT TCGTATTCAT AACTTTACTA	900
TTAGGTGCCT ATTTTCTACT TTTGTCAPTA CTCACCTTGC TGACTTTATT AAAACATAAA	960
AATGGATTTT TAATTGTGAT GCCCTTTATT TTATTTTCCA TTCACCTTGC TTATGGCCTT	1020
GGGACGATTG TAGGTTTAAT TAGAGGATTT AAATGGAAGA AGGAGTACAA GAGAACAATA	1080
ATTTATTTGG ATAAAATAAG CCAAATAAAT CAAATATGC TATAATAACA ATATAGTAAA	1140
ACTCTTTTAA GGAGGAGTAG ATTTCTATGA ATAAAAAACT AACAGATTAT GTGATTGATC	1200
TGGTGGAAT TTTAAATAAA CAACAAAAGC AGGTTTCTG GGAATATTT GATATTTTCA	1260
GTATGGTGGT TTCCATCATT GTATCTTATA TTTTATTTTA TGGGCTGATT AATCCAGCAC	1320
CTGTTGACTA CATTATCTAT ACGAGTTTGG CCTTCCTGTT CTATCAATTG ATGATTGGTT	1380
TTTGGGGGTT GAACGCGAGC ATTAGTCGTT ACAGCAAGAT TACGGATTTT ATGAAAATCT	1440
TTTTTGGTGT GACTGCTAGC AGTGTCTTGT CATATAGTAT CTGTTATGCC TTCTTGCCAC	1500
TCTTCTCCAT CCGTTTCATC ATTCTCTTTA TCTTGTGAG TACCTTCTTG ATTTTATTGC	1560

1050

CACGGATTAC TTGGCAGTTA ATCTACTCCA GACGCAAAAA AGGTAGTGGT GATGGAGAAC	1620
ACCGTCGGAC CTTCTTGATT GGTGCCGGTG ATGGTGGGGC TCTTTTATG GATAGTTACC	1680
AACATCCAAC CAGTGAATTA GAACTGGTCG GTATTTTGGG TAAGGATTCT AAGAAAAAGG	1740
GTCAAAAAC TGGTGGTATT CCTGTTTGGG GCTCTTATGA CAATCTGCCT GAATTAGCCA	1800
AACGCCATCA AATCGAGCGT GTCATCGTTG CGATTCCGTC GCTGGATCCG TCAGAATATG	1860
AGCGTATCTT GCAGATGTGT AATAAGCTGG GTGTCAAATG TTACAAGATG CCTAAGGTTG	1920
AAACTGTTGT TCAGGCCTT CACCAAGCAG GTACTGGCTT CCAAAAAATT GATATTACGG	1980
ACCTTTTGGG TCGTCAGGAA ATCCGTCTTG ACGAATCGCG TCTGGGTGCA GAACTGACAG	2040
GTAAGACCAT CTTAGTCACA GGAGCTGGAG GTTCAATCGG TTCTGAAATC TGTCGTCAAG	2100
TTAGTCGCTT CAATCCTGAA CGCATTGTCT TGCTCGGTCA TGGGGAAAAC TCAATCTACC	2160
TTGTTTATCA TGAATTGATT CGTAAGTTCC AAGGGATTGA TTATGTACCT GTGATTGCGG	2220
ACATTCAAGA CTATGATCGT TTGTTGCAAG TCTTTGAGCA GTACAAACCT GCTATTGTTT	2280
ATCATGCGGC AGCCCAACAG CATGTTCTTA TGATGGAGCG CAATCCAAAA GAAGCCTTCA	2340
AAAACAATAT CCGTGGAAC TACAATGTTG CTAAGGCTGT TGATGAAGCT AAAGTGTCTA	2400
AGATGGTTAT GATTTGACA GATAAGGCAG TCAATCCACC AAATGTTATG GGAGCAACCA	2460
AGCGCGTGGC GGAGTTGATT GTCACTGGCT TTAACCAACG TAGCCAATCA ACCTACTGTG	2520
CAGTTGTTTT TGGGAATGTT CTTGGTAGCC GTGGTAGTGT CATCCAGTC TTTGAACGTC	2580
AGATTGCTGA AGTGCGCCT GTAACGGTGA CAGACTTCCG TATGACCCGT TACTTTATGA	2640
CCATTCCAGA AGCTAGCCGT CTGGTTATCC ATGCTGGTGC TTATGCCAAA GATGGGGAAG	2700
TCTTTATCCT TGATATGGGC AAACCAGTCA AGATTTATGA CTTGGCCAAG AAGATGGTGC	2760
TTCTAAGTGG CCACACTGAA AGTGAAATTC CAATCGTTGA AGTTGGAATC CGCCCAAGTG	2820
AAAACTCTA CGAAGAACTC TTGGTATCAA CCGAACTCGT TGATAATCAA GTTATGGATA	2880
AGATTTTCGT TGGTAAGGTT AATGTCATGC CTTAGAATC CATCAATCAA AAGATTGGAG	2940
AGTTCGCAC TCTCAGTGA GATGAGTTGA AGCAAGCTAT TATCGCCTTT GCTAATCAAA	3000
CAACCCACAT TGAATAAAAA AGAAAAACGC ATAGTATCAA GTTACACAAC CTTGGTAATA	3060
TGCGTTTTAT TATGTAGAGA CTTATACTCT TCGAAAATCT CTTCAAACCA CGTCAACGTC	3120
GCCTTGCCGT ATATGGTTAC TGAATTCGTC AGTTCTATCC ACAACCTCAA AACAGTGTTC	3180
TGAGTGGACT TCGTCAGTTC TATCCACAAC CTCAAACAG TGTTTTGAGC TGACTTCGTC	3240
AGTTCTATCC ACAACCTCAA AACAGTGTTC TGAGCTGAGT TCGTCAGTTC CATCCACAAC	3300
CTTAAACAG TGTTTTGAGT TGACTTTCGT CAGTTCCATC TACAACCTTA AAACAGTGT	3360

1051

TTGAGCTGCC CGCAGCTAGT TTCCTAGTTT GCTCTTTGAT TTTCATTGAG TATTACTTCA	3420
TTTTCTTCTG AAATGGAATT GTTACCCAGT CTATGCTATT GAAAATACGC CAAAACCTCT	3480
AAGGGTTTGT GAGCGATATA ATCAGGTTGA TAGTTTAGTA GATCTGCTTG CTCTCCAAAT	3540
CCCCAAGTGA TGGCCAATTT CTGAATACCT GTTCTCGAG CTCCCAGCAT ATCAAACCTG	3600
GTATCTCCGA TGATGATGGC TTGTTCTGGT GCTAGTTGAT GTGTCTGCAA GGCTTGGTGA	3660
ATGACATCTG CTTATGGGG TGCTTCAGGG CTAGAACCAT AAATGCCATC AAAGAAATGA	3720
TGGATTTCOA AGTTTTTTCG CATGTCTTGA GCAGTAGATG TATCCTTTGT CGTGGTGATG	3780
TAGAGTGGAT AACTGCTCGA TAACTCCTCA AGCAAGTCTA TAATCTGAGG AAAGAGTTGA	3840
GCTTCATAGA TGCTTTTTCG CTTATAGTAA GAACGATATA TCTGCACGGC TTCAGAAATT	3900
TGGTCTTTGG ACAGGCAGGT CGCAAACTA CTTTCGAGAG GTGGTCCCAT AAAACCACGA	3960
ATAGTTTTGG CATCAGGGCT AGGCACCCCC AGCTCTTTAA AGGTATAGGT AAAGGCATTG	4020
TGAATCCCGA TAGAACTATC AACGAGGGTT CCATCCAAAT CGAAAAAAT CGCTGTGATA	4080
GAGGTCATGG TTTCTCCTAT TTGATAAGCT TATTCTCCGA AAATTTCTTT TTGGAGGCGA	4140
CGACCAGTAG GGGTGGTAGC GAGTCCACCT TCAGCTGTTT CACGAAAGGC AGTTGGCATG	4200
CTTGCTCCTA CTTGGTACAT GGCATCGATC ACTTCATCCA CAGGGATTTT AGATTTCGATA	4260
CCTGCCAAGG CCATGTCTGC TGCGATGAAA GCAAAGCTAG CTCCCATGGC ATTACGTTTG	4320
ACACAGGGAA CTTGACCAA ACCTGCAACA GGGTCACAGA TGAGGCCTAG CATATTTTTA	4380
ATGACAAAGG CAATAGCTTG ACTGGCCTGA TAAGGTGTTT CACCTGCAGC CAGAGTCAAG	4440
GCGGCAGCAC TCATAGCAGA GGCTGAACCA ACTTCAGCTT GACACCCACC CTCAGCACCT	4500
GAGATGGAGG CATTGTTTGC GATGACTAGT CCAAAGGCAC CAGCAGCAA GAGGAAATCC	4560
AATTGTTGCT CGTGGCTGAG GTCTAATTTT TCAATAGCAG CAGTGAGAAC GGATGGCAGA	4620
CAGCCAGCAC TTCCAGCGGT TGGAGTGGCA CAGACCAAGC CCATTTTGGC ATTGTGTTCA	4680
TTGACTGCCA TGGCATTTTC GGCAGCCGAG AGAATCGTAT AATCTGACAG AGTTTTTCCG	4740
TTTTCGATGT AGTGATCCAA TTTGGCAGCA TCTCCACCTG TCAGGCCACT ACGAGATTTA	4800
TTTTCATGA GGCCAAGTTG GACAGAGGCT TTCATACTT CCAGATTGCG TTCCATGAGA	4860
AGGAAGACTT CTTACGTTT CCGACCGGTC AATTCAACT CTGTTGTAAT CATGAGTTCT	4920
GCGACATTTT CTTGAAAGTC CAGATCTGCT TGCTCGACCA ATTCTTTGAT AGAATAAAAC	4980
ATGCTTCCTC CTATTTAAAG AAATTGACAT TGTGGAGATG AGGGATTTT CGAATTTCTT	5040
CGATAGCCTC ATCAGATTG CCACTGTCAA CTTGATAAT CATAATGGCT TTTTCACCAG	5100

1052

CTTTTTCACG AGTGACATTC ATCTGGGCGA TATTGATACC ATAGCGGGAA AGCGCCTCTG	5160
TAACAAGGGC AATCATACCT GGAATATCTT GATGAACGAT GATGATAGTC GGTGTATTCA	5220
TATTGAGAGA GACGGCAAAA CCATTGAGTT CGGTACCTG AATATTTCTT CCACCGATAG	5280
AAATACCACT CACGCTGATG GTCTTGTTGG CATTTTAAAC AGTAATTTTA GTGGTGTTAG	5340
GGTGAGGGGC ATTGCTGTCT TTCTGAATGG TCCAGACAAT CTTGATACCA CGCTTGTTGG	5400
CAATTTCCAG ACTATTTGGA ATTTTCAGGAT CATCTGTATC CATTCTCTAA ATACCTGCAA	5460
CAAGGGCTAG GTCTGTTCCG TGACCACGAT AGGTCTTGGC AAATGAGTTA AAAAGTTGGA	5520
ATTCAACTTC TGTCGGAGTA TCATCAAAAA TGAAGAGAC AATCTTCCCA ATACGAACAG	5580
CACCAGCGGT ATGGCTACTA GATGGGCCAA TCATAACTGG TCCGATGATA TCAAAGACAG	5640
ATTGAAAACG AAGTGATTTC ATCAGTTTCC CTTTATAAAA ATTCTTATCT CTATTATATC	5700
AAAGAATGAG GGGCTTGGCT TTAATTGTGG ATGAAAACCT TTCTAATACC TCAAATAGCA	5760
TAAAAATAGT ATCTTTTATG ACAAAAAACA CTTATTTAG GGAAATAAAA AATAATTTTG	5820
TAATATTTCT ACATAAAAGT GTCAAGAAAC GGTAATATTT AAAGGGTATG ATAGAACTAT	5880
AGAAAGAAGG AGAATTTTCG AATATGAAAT CAATACTAA AAAGATTTAA GCAACTCTTG	5940
CAGGAGTAGC TGCCTTGTTT GCAGTATTTG CTCCATCATT TGTATCTGCT CAAGAATCAT	6000
CAACTTACAC TGTTAAAGAA GGTGATACAC TTTCAGAAAT CGCTGAAACT CACAACACAA	6060
CAGTTGAAAA ATTGGCAGAA AACAACCACA TTGATAACAT TCATTTGATT TATGTTGATC	6120
AAGAGTTGGT TATCGATGGC CCTGTAGCGC CTGTTGCAAC ACCAGCGCCA GCTACTTATG	6180
CGGCACCAGC CGCTCAAGAT GAACTGTTT CAGCTCCAGT AGCAGAAACT CCAGTAGTAA	6240
GTGAAACAGT TGTTTCAACT GTAAGCGGAT CTGAAGCAGA AGCCAAAGAA TGGATCGCTC	6300
AAAAAGAATC AGGTGGTAGT ATACAGCTAC AAATGGACGT TATATCGGAC GTTACCAATT	6360
AACAGATTCA TACCTGAACG GTGACTACTC AGCTGAAAAAC CAAGAACGGG TACCG	6415

## (2) INFORMATION FOR SEQ ID NO: 163:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8494 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

TACCCCTTTC GAATTTTGGC AAAAATTCGG TAAGGCTTTG ATGGTAGTTA TCGCGTTAT	60
CCCGGCTGCT GGTGATGA TTTCAATCGG TAAGTCTATC GTGATGATTA ACCCAACCTT	120

1053

TGCACCACTT GTCATCACAG GTGGAATTCT TGAGCAAATC GGTGGGGGG TTATCGGTAA	180
CCTTCACATT TTGTTTCCCC TAGCCATTGG AGGAAGCTGG GCTAAAGAAC GTGCTGGTGG	240
TGCTTTCGCC GCTGGTCTTG CCTTCATCTT GATTAACCGT ATCACTGGTA CAATCTTTGG	300
TGTATCAGGC GATATGTTGA AAAATCCAGA TGCTATGGTA ACTACTTTCT TTGGTGGTTC	360
AATCAAAGTT GCTGATTACT TTATCAGTGT TCTTGAAGCT CCAGCCTTGA ACATGGGGGT	420
ATTCGTAGGG ATTATCTCAG GTTTGTAGG GGCAACTGCT TACAACAAAT ACTACAACCTT	480
CCGTAAACTT CCTGATGCAC TTTCATTCTT CAACGGGAAA CGTTTCGTAC CATTTGTAGT	540
TATTCTTCGT TCAGCAATCG CTGCAATTCT ACTTGCTGCT TTCTGGCCAG TAGTTCAAAC	600
AGGTATCAAT AACITCGGTA TCTGGATTGC CAACTCACAA GAAACTGCTC CAATTCTTGC	660
ACCATTCTTG TATGGTACTT TGGAACGTTT GCTCTTGCCA TTGGTCTTC ACCACATGTT	720
GACTATCCCA ATGAACTACA CAGCTCTTGG TGGTACTTAT GACATTTTAA CTGGTGCAGC	780
TAAAGTACT CAAGTATTCG GTCAAGACCC ACTATGGCTT GCATGGGTAA CAGACCTGT	840
AAACCTTAAA GGTACTGATG CTAGTCAATA TCAACACTTG TTAGATACAG TACATCCAGC	900
TCGTTTCAAA GTTGACAAA TGATCGGTTT ATTCGGTATC TTGATGGGTG TGATTGTTGC	960
TATCTACCGT AATGTTGATG CTGACAAGAA ACATAAATAC AAAGGTATGA TGATTGCAAC	1020
AGCTCTTGCA ACATTCTTGA CAGGGGTAC TGAACCAATC GAATACATGT TCATGTTTAT	1080
CGCAACACCT ATGTATCTTG TTTACTCACT TGTTCAGGT GCTGCCTTCG CTATGGCTGA	1140
CGTCGTAAAC CTACGTATGC ACTCATTCCG TTCAATCCAG TTCTTGACTC GTACACCTAT	1200
TGCAATCAGT GCTGGTATTG GTATGGATAT CGTTAACTTC GTTTGGGTAA CTGTTCTCTT	1260
TGCTGTAATC ATGTACTTTA TCGCAAACCT CATGATTCAA AAATCAACT ACGCAACTCC	1320
AGGGCGCAAC GGAAACTACG AAACTGCTGA AGGTCAGAA GAAACCAGCA GCGAAGTGAA	1380
AGTTGCAGCA GGCTCTCAAG CTGTAAACAT TATCAACCTT CTTGGTGGAC GTGTAAACAT	1440
CGTTGATGTT GATGCATGTA TGAATCGTCT TCGTGTAATC GTTAAAGATG CAGATAAAGT	1500
AGGAAATGCA GAGCAATGGA AAGCAGAAGG AGCTATGGGT CTTGTCTATGA AAGGACAAGG	1560
GGTTCAAGCT ATCTACGGTC CAAAAGCTGA CATTTTGAAG TCTGATATCC AAGATATCCT	1620
TGATTCAAGT GAAATCATTC CTGAAACTCT TCCAAGCCAA ATGACTGAAG CACAACAAAA	1680
CACTGTTTCA TTCAAAGATC TTAATGAGGA AGTTTACTCA GTAGCAGACG GTCAAGTTGT	1740
TGCTTTGGAA CAAGTAAAGG ATCCAGTATT TGCTCAAAAA ATGATCGGTG ATGGATTTGC	1800
AGTAGAACCT GCAAAATGAA ACATTGTATC TCCAGTTTCA GGTACTGTGT CAAGCATCTT	1860

1054

CCCAACAAAA	CATGCTTTTG	GTATTGTGAC	GGAAGCAGGT	CTTGAAGTAT	TGGTTCACAT	1920
TGGTTTGGAC	ACAGTAAGTC	TTGAAGGTAA	ACCATTTACA	GTTTCATGTTG	CTGAAGGACA	1980
AAAAGTTGCA	GCAGGAGATC	TCCTTGTCAC	AGCTGACTTG	GATGCTATCC	GTGCAGCAGG	2040
ACGTGAAACT	TCAACAGTAG	TTGTCTTCAC	AAATGGTGAT	GCAATTAAAT	CAGTTAAGTT	2100
AGAAAAACA	GGTTCTCTTG	CAGCTAAAAC	AGCAGTTGCT	AAAGTAGAAT	TGTAATATAC	2160
TTGAGGTTGG	AAGCTGTATT	CCAACCTCTT	ATTTTGGGAG	AAAAGAATGA	AATTTTAAAC	2220
ACTCAATACT	CACAGTTGGA	TGGAGAAAGA	AGCAGAGGAA	AAATTCAGGA	TTTTGCTTGA	2280
AGATATTCTT	GAAAAGGACT	ATGATTTGAT	TTGTTTTCAA	GAAATCAATC	AGGAGATGAC	2340
CTCGTCAGAG	GTGGAGGTTA	ATGACCTTTA	TCAAGCTTTG	CCAGCAGCTG	AGCCTATTCA	2400
CCAAGACCAT	TATGTTAGAC	TCTTGGTTGA	AAAGTTGTCT	GAGCAAGGGA	AAAATTACTA	2460
CTGGACCTGG	GCCTATAACC	ATATCGGCTA	TAACCGCTAC	CACGAAGGTG	TGGCTATCTT	2520
GTCTAAAACA	CCTATTGAAG	CCAGAGAAAT	TTTGGTTTCA	GATGTGGATG	ATCCAACAGA	2580
CTATCATACT	CGCCGTGTTG	CCCTAGCTGA	AACTGTAGTC	GATGGCAAGG	AGCTAGCAGT	2640
TGCCAGTGTT	CATCTCTCTT	GGTGGGATAA	AGGTTTCCAA	GAAGAATGGG	CACGATTTGA	2700
GGCTGTCTTG	AAAAAATTGA	ACAAGCCACT	TTTACTAGCT	GGAGATTTCA	ACAATCCGGC	2760
TGGACAGGAA	GGTTACCAAG	CTATTTTAGC	TAGTCCATTA	GGCTTACAAG	ACGCATTTGA	2820
AGTTGCTCAA	GAGAAAAGTG	GTAGCTATAC	TGTTCCGCCT	GAAATTGATG	GCTGGAAAGG	2880
GAACACTGAA	CCCTTCGAA	TCGATTATGT	CTTTACTACC	AAAGAGTTAG	CGGTGGAAAA	2940
TTTACATGTC	GTATTTGATG	GTAACAAGAG	TCCACAAGTG	AGTGATCACT	ATGGCTTGAA	3000
TGCTATATTA	AACTGGAAAT	AATAACTGAA	AAGAGTTTGG	AACTATAAAA	TTCCAGCCTT	3060
TTCTTACTAG	AGAAGCTACT	GGAAATAGCC	TAAATAAGTG	AGACTACTGT	AATGGAATAA	3120
AATATGGTAT	AATTGATAAG	GTAGATAGAA	TCGAGGATGT	TATGTCATTT	ACGAAATTTC	3180
AATTTAAAAA	CTATATTAGA	GAAGCCTTGA	AGGAGTTAAA	ATTTACAAC	CCAACAGAGG	3240
TGCAAGACAA	GTTGATTCCT	ATTGTTTTGG	CAGGTCGTGA	CCTAGTAGGA	GAATCAAAAA	3300
CAGGTTCAGG	TAAGACTCAT	ACTTCTTGT	TACCGATTTT	CCAGCAATTA	GATGAAGCTA	3360
GCGATAGTGT	ACAAGCAGTG	ATTACTGCAC	CGAGTCGTGA	GTTGGCTACT	CRAATTTACC	3420
AAGTAGCGCG	TCAGATTTCA	GCTCACTCAG	ATGTCGAAGT	TCGTGTGGTT	AATTATGTGG	3480
GTGGTACGGA	TAAGGCTCGC	CAGATTGAGA	AATTGGCAAG	CAATCAGCCT	CATATTGTTA	3540
TTGGAACACC	AGGCCGTATC	TACGACTTGG	TTAAATCTGG	TGATTTAGCT	ATTCATAAAG	3600
CCAAGACATT	TGTTGTTGAT	GAAGCAGATA	TGACCTTGGA	TATGGGATTC	TTGGAAACTG	3660

1055

TTGATAAGAT TGCTGGCAGT CTTCCAAAAG ACTTGCAATT CATGGTCTTC TCAGCGACTA	3720
TCCCACAAAA ACTGCAACCA TTCTTGAAAA AATACTTATC AAATCCTGTT ATGGAGAAAA	3780
TTAAGACCAA AACGGTTATT TCTGACACCA TTGATAATTG GTTGATTTCG ACCAAGGGAC	3840
ATGATAAGAA TGCTCAAATT TACCAGTTGA CTCAGTTGAT GCAGCCGTAT TTGGCAATGA	3900
TTTTTGTTAA CACTAAAACG CGTGCTGATG AATTGCATTG ATATCTGACT GCTCAAGGCT	3960
<hr/>	
TGAAGGTTGC AAAAATCCAT GCGGATATTG CCCCTCGTGA ACGCAAGCGA ATCATGAATC	4020
AGGTGCAAAA TCTGGATTTT GAGTATATTG TCGCAACAGA TTTGGCAGCG CGTGGGATTG	4080
ACATTGAAGG TGTCAGCCAT GTCATCAATG ATGCCATTCC GCAAGACTTA TCTTTTPTTG	4140
TTTCATCGTGT TGGTCGTACT GGACGAAATG GCCTACCAGG TACAGCTATT ACCCTTTATC	4200
AGCCAAGTGA TGACTCGGAT ATCCGTGAGT TGGAGAAATT GGGAATCAAG TTTAGTCCTA	4260
AGATGGTCAA AGACGGGGAA TTTCAAGATA CCTATGACCG TGATCGTCGT GCCAACCGTG	4320
AGAAAAACA AGATAAACTT GATATCGAAA TGATTGGTTT GGTTAAAAAG AAAAGAAAA	4380
AAGTCAAACC GGGTTATAAG AAGAAAATTC AATGGGCGGT TGATGAAAAG CGCCGTAAAA	4440
CCAAGCGTGC TGAATTCGC GCTCGCGGTC GTGCAGAGCG TAAAGCTAAA CGCCAAACAT	4500
TTTAATAGAA ATTGTTGGAG TATTGAGCTC CAACTTTTTT ATTTATGAGA ACGAACTATC	4560
TAAACCGAAA CACTACATTA AAGACTGCAA ATTGCGATTA AAAATGGTAT AATGATAAAG	4620
TTATATAGTC CCGATAAGAT GGTAGGTATT TATTACGAAG AGTTTTCTTA TCAGTACTTT	4680
GTAACCTCTAT AACAATATTT TTTAAGGGGG GACATTTTTA TGTCAGAGCG TAAATTATTC	4740
ACGTCTGAAT CTGTATCTGA GGGGCATCCG GATAAGATTG CAGACCAAAT TTCAGATGCG	4800
ATTTTGGATG CTATTTTAGC AAAGGATCCA GAGGCGCAGC TTGCTGCTGA AACAGCTGTA	4860
TATACTGGTT CTGTCCACGT TTTTGGTGAA ATTTCTACAA ATGCCTATGT GGATATTAAC	4920
CGTGTGGTTC GTGATACCAT TGCAGAGATT GGTATACCA ATACAGAATA TGGATTTTCT	4980
GCTGAGACGG TGGGAGTACA CCCATCTTTG GTGGAACAAT CTCCTGACAT CGCTCAAGGT	5040
GTTAACGAAG CCTTGGAGGT TCGTGGAAT GCTGATCAAG ATCCACTGGA CTTGATTGGA	5100
GCAGGTGACC AAGGGCTCAT GTTTGGATTT GCAGTAGATG AAACAGAAGA GCTTATGCCA	5160
TTGCCAATTG CACTCAGTCA TAAATTGGTT CGTCGTCTGG CAGAACTTCG TAAGTCTGGA	5220
GAAATTAGCT ATCTCCGTCC AGATGCAAAA TCACAAGTTA CAGTTGAGTA CGATGAAAAT	5280
GACCGTCCGG TACGTGTAGA TACAGTCGTT ATTTCTACTC AGCATGATCC AGAGGCCACT	5340
AATGAACAAA TCCATCAAGA TGTGATTGAC AAGGTCATCA AAGAAGTTAT TCCATCTTCT	5400

1056

TATCTTGATG	ATAAGACAAA	ATTCTTTATC	AATCCGACAG	GTCTTTTGT	AATCGGTGGT	5460
CCTCAAGGGG	ACTCAGGTTT	GACTGGTCGT	AAGATTATTG	TAGATACTTA	TGGTGGCTAC	5520
TCTCGTCATG	GTGGTGGTGC	CTTCTCTGGT	AAAGATGCCA	CTAAGGTGGA	TCGTTCAGCC	5580
TCTTATGCGG	CTCGCTATAT	TGCCAAGAAT	ATCGTTGCAG	CAGACCTTGC	TAAGAAGGCA	5640
GAAGTGCAGT	TGGCCTATGC	TATCGGTGTT	GCGCAACCTG	TTTCTGTTCT	TATCGATACT	5700
TTCCGTACAG	GAACAGTAGC	TGAAAGTCAA	CTTGAAAAAG	CGGCTCGTCA	AATCTTTGAC	5760
CTTCGCCCTG	CAGGGATTAT	CCAAATGCTG	GACCTCAAGC	GTCCAATTTA	CCGTCAAACA	5820
TCGGCTTACG	GTCACATGGG	ACGTACAGAT	ATTGATCTTC	CATGGGAACG	TTTGGATAAG	5880
GTAGATGCTT	TGAAAGAAGC	AGTAAAATAA	GATTTTAAGA	GGGGAACGTC	CTCTCTTTT	5940
TATAGTTTTT	AACTATACTG	GGATACTGTT	CTGAAAATCC	ATTTTGCGAA	AGTAGAGATT	6000
TACATGTATA	GTAGATTGAA	ACTAGAATAG	TACACCTCAA	CTTCTAAAAC	ATTGTTAGCA	6060
ATCAATTTGA	CTGTCCTGAT	CGATTTCTCC	TGTTCTTGTT	TCATTTTACT	ATATTTCTTT	6120
AAAAATGATA	AAGGTTAAGA	TTTCTCCTCG	TAATAGATAA	TCTTGGGGAT	ATTTCAATCC	6180
AAAGTTTTAT	TCGTTATCAC	TTGACTATTG	CAAGGTTTTC	TAGAGCAACA	GAGTCATGGA	6240
ATGGACTCAT	GGTTGAGATT	TCTCCTTGTT	GCTTGGACTT	CATTCAAAAG	TCTGTTACCC	6300
AAGCCTTGTT	CAAACTTCTA	ATACACTAGC	TGTTTCCATA	GCATGACTTC	TGTACTAGAC	6360
TTTCTTTTCC	GAATAAATAG	ATAGAACCAC	AGAATCTAGT	AAACCTAGAA	TTAAAATTAT	6420
GGTATAATAT	TAGCAATAAA	AGAAATCTGG	AGGATTAGAA	TCATGGTATC	AACGAAAACA	6480
CAAATTGCTG	GTTTTGAGTT	TGACAATTGC	TTGATGAATG	CAGCAGGTGT	GGCTTGATG	6540
ACGATAGAGG	AGTTAGAAGA	GGTCAAAAAC	TCAGCGGCAG	GAACCTTTGT	TACTAAGACA	6600
GCGACCTTGG	ACTTCCGTCA	GGGGAATCCT	GAGCCACGCT	ACCAAGATGT	TCCACTTGGT	6660
TCCATCAACT	CTATGGGCTT	GCCAAATAAT	GGCTTAGACT	ATTATTTGGA	TTATCTTTTA	6720
GATTTGCAGG	AAAAAGAGTC	GAACCGAACT	TTCTTCTTAT	CTCTGGTCGG	CATGTCTCCA	6780
GAGGAAACCC	ATACTATTTT	GA AAAAAGTC	CAAGAGAGTG	ATTTTCGTGG	TCTGACTGAG	6840
CTAAATCTTT	CCTGTCCAAA	TGTTCCAGGT	AAACCTCAGA	TTGCCTATGA	TTTTGAGACA	6900
ACAGACCGGA	TTTTGGCAGA	AGTGTTCGCT	TACTTCACCA	AACCTCTTGG	AATTAAATTG	6960
CCACCTTATT	TTGATATTGT	TCACCTTGAC	CAAGCGGCAG	CTATTTTCAA	CAAATATCCG	7020
CTCAAGTTTG	TCAACTGCGT	TAACTCTATC	GGAAACGGCC	TCTATATAGA	AGACGAATCT	7080
GTCGTTATT	GGCCTAAGAA	TGGTTTTGGT	GGAATTGGTG	GAGAATACAT	CAAACCGACT	7140
GCTTTAGCCA	ATGTTACGCG	CTTTTATCAA	CGTTTAAATC	CTCAAATCCA	AATTATCGGA	7200



1057

ACAGGTGGCG TTCTGACTGG TCGAGATGCC TTTGAACACA TCCTCTGTGG AGCAAGTATG	7260
GTGCAGGTGG GAACGACCCT TCACAAAGAA GCGCTCAGTG CTTTGTACCG CATTACCAAT	7320
GAAGTAAAAG CAATCATGGT GGAAAAAGGC TACGAGAGCT TAGAAGATTT CCGTGGGAAA	7380
TTGCGCTATA TTGACTAAAT TAAATCGAAA AATCTGAAGA AAGGAGAGAC GATGCTAGCC	7440
ATTGAAGAAA GTCAGAAGTT GACTTTATCA AATTTACCGA GCCTGAGCCT ATTTACAGGG	7500
ACAGATCAGG GTCAGTTTGA AGTGATGAAG AGTCAAATGT TGAAACAGAT TGGGTATGAT	7560
TCTGCTGACC TCAACTTTGC CTACTTTGAT ATGAAAGAAG TAGTTTACAA GGATGTGGAA	7620
CTGGAGTTGG TCAGCCTTCC TTTCTTTGCG GATGAGAAAA TCGTGATATT AGATTATTTT	7680
ATGGATATCA CGACTGCTAA GAAACGCTTT TTGACAGATG ATGAGCTTAA GTCATTGAG	7740
GAATACCTTG ACAATCCTTC TCCAACAACC AAGTTGATAA TCTTTGCAGA AGGAAAGCTG	7800
GATAGCAAAA GACGGTTACT CAAATTACTT AAGCGTGATG CCAAGGCCTT CGATGCAGTA	7860
GAAGTAAAAG AACAAGAATT GCGCCAGTAC TTCCAAAAGT GGAGTCAGAA ACAAGGTCTG	7920
CAGTTTACCA ATCATTCTTT TGAAAATCTC CTCATCAAGT CGGGGTTTCA ATTTAGCGAA	7980
ATCCAGAAAA ATCTTCTCTT TTTACAGTCC TATAAGGCGA ATTCTGTTAT TGAGGAAGAG	8040
GATATTGTTA ACGCAATTCC CAAGACTTGC AGGACAATAT TTTTGATTTA ACTCAGTTTA	8100
TTCTGACTAA AAAGATGGAT CAGGCGCGCG ATTTGGTGAG AGACTTGACC TTGCAAGGGG	8160
AAGATGAAAT CAAACTGATT GCAGTCATGC TGGGACAATT TCGGACTTTT ACTCAGGTGA	8220
AGATTTTGGC GGAGTCTGGC CAAACAGAAT CGCAGATTGC AAGTAGTTTA GGTAGTTATC	8280
TGGGACGTAA CCCAAATCCT TATCAAATCA AGTTTGCATT AAGAGATTCTG AGAGGACTTT	8340
CTTTGAGCTT TTTGAAGCAA GCTATTCCT ATTTGATTGA GACAGACTAT CAGATTAAGA	8400
CAGGTCTTTA TGAAAAAGGT TTCCTTTTGG AAAAGGCACT CTTACAGATT GCTAGTCAGG	8460
TCAATTGACA TTTGTTGAAA CTACTAACCC GCGG	8494

(2) INFORMATION FOR SEQ ID NO: 164:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9707 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

CCGGTCAGTT CGTTCAGTAC AAGGAATCAT AATGAACGAT CAATCAGAAA AAAAGACTAG 60

1058

AAAGAAGACT GTATGGATAA TCGACCAATT GGTTTTTTGG ATTCGGGTGT CGGGGGCTTG	120
ACCGTTGTGC GCGAGCTCAT GCGCCAGCTT CCCCATGAAG AAATCGTCTA TATTGGAGAT	180
TCGGCGCGGG CGCCCTATGG CCCCCGTCCT GCTGAGCAAA TTCGTGAATA TACTTGCCAG	240
CTGGTCAACT TTCTCTTGAC CAAGGATGTC AAAATGATTG TCATTGCTTG TAACACTGCC	300
ACTGCGGTCTG TCTGGGAAGA AATCAAGGCT CAACTAGATA TTCTGTCTT GGGTGTAATT	360
TTGCCAGGAG CTTCCGCAGC CATCAAGTCC AGTCAAGGTG GGAAAATCGG AGTGATTGGA	420
ACGCCCCATGA CGGTACAATC AGACATATAC CGTCAGAAAA TCCATGATCT GGATCCCGAC	480
TTACAGGTGG AGAGCTTGGC CTGTCCCAAG TTTGCTCCCT TGGTTGAGTC AGGTGCCCTG	540
TCAACCAGTG TTACCAAGAA GGTGGTCTAT GAAACCTGC GTCCCTTGGT TGGAAAGGTG	600
GATAGCCTGA TTTTGGGCTG TACTCATTAT CCACTCCTTC GCCCTATTAT CCAAAATGTG	660
ATGGGGCCAA AGGTCAGCT CATCGATAGT GGGGCAGAGT GCGTACGGGA TATCTCAGTC	720
TTACTCAATT ATTTTGAAAT CAATCGTGGT CGCGATGCTG GACCACTCCA TCACCGTTTT	780
TACACAACAG CCAGTAGCCA AAGTTTGTCA CAAATTGGTG AAGAATGGCT GGAAAAAGAG	840
ATTCATGTGG AGCATGTAGA ATTATGACAA ATAAAAATTA TGAATATAAG GATGACCAGG	900
ACTGGTATGT TGGGTCTTAT AGTATTTTGT GTGGCGTTAA CAGTTTGAGC GACTATAAGA	960
CAGATTTTCC TCTGTTTGAA TTCTCCAAAA TATTTGGAGA TGAAGAGTAT GGTTTCCCGC	1020
TTTCAGTTAC TGTTTTACGC TATGGTTCTA TCTACCGTTT GTTCTCCTTT GTGGTAGACA	1080
TGCTTAATCA AGAAATGGGA CGAAACTTGG AAGTTATTCA ACGTCATGGG GCCCTGCTCT	1140
TGGTTGAAAA TGGGCAACTC TTGTATGTAG AATTGCCTAA AGAAGGGGTC AATGTTTCATG	1200
ATTTCTTTGA GACAAGCAAG GTCAGAGAAA CCTTGTGAT TGCGACTCGT AACGAAGGTA	1260
AAACCAAGGA ATTCCGAGCT ATCTTTGATA AGTTAGGCTA CGATGTGGAA AATCTTAATG	1320
ACTACCCTGA CCTGCCTGAA GTAGCAGAAA CAGGTATGAC CTTTGAAGAA AATGCCCGCC	1380
TTAAGGCAGA AACCATTCT CAATTAACGG GCAAGATGGT TTTGGCAGAT GATTCTGGTC	1440
TCAAAGTCGA TGTCTTGGT GGCTTACCAG GCGTCTGGTC AGCTCGTTTC GCAGGTGTGG	1500
GAGCAACTGA CCGTGAAAAA AATGCCAAAC TCTTGACGA ATTGGCCATG GTCTTTGAAC	1560
TCAAGGACCG CTCGGCTCAG TTCCACACAA CCCTAGTCGT AGCCAGCCCA AATAAGGAAA	1620
GTTTAGTTGT TGAAGCAGAC TGGTCAGGTT ATATTAACTT TGAACCTAAG GGTGAAAATG	1680
GCTTTGGCTA TGATCCCCTC TTCCTGTAG GAGAAACAGG TGAGTCATCA GCTGAATTAA	1740
CCCTGGAAGA AAAAAATAGT CAATCTCACC GTGCCTTAGC CGTTAAGAAA CTTTGGAGG	1800
TATTTCCATC ATGGCAAAGC AAACCATCAT TGTAATGAGC GATTCCCATG GCGATAGCTT	1860

1059

GATTGTGGAA GAAGTCCGTG ATCGCTATGT GGGCAAAGTC GATGCTGTTT TTCATAACGG	1920
CGATTCTGAA CTACGTCCGG ATTCTCCACT TTGGGAGGGC ATCCGCGTTG TTAAGGGAA	1980
CATGGACTTC TACGCCGGCT ACCCAGAACG TCTGGTGACT GAGCTTGGTT CGACCAAGAT	2040
TATCCAAACT CATGGTCACT TGTTTGACAT CAATTTCAAC TTTCAAAAGT TGGACTACTG	2100
GGCTCAGGAG GAAGAGGCCG CTATCTGCCT CTATGGTCAC TTGCATGTGC CAAGTGCTTG	2160
GTTGGAAGGC AAGATCCTCT TTCTAAATCC AGGTTCTATC AGTCAACCAC GAGGTACCAT	2220
CAGAGAATGT CTCTATGCTC GTGTGGAGAT TGATGATAGT TACTTCAAAG TGGACTTTTT	2280
GACACGAGAT CACGAGGTGT ATCCAGGTTT GTCCAAGGAG TTTAGCCGAT GATTGCCAAG	2340
GAGTTTGAGA CTTTCTTGTT GGGGCAGGAG GAAACTTTTT TGACCCCTGC TAAAACTA	2400
GCTGTGTTGA TTGATACCCA CAATGCGGAT CATGCGACCC TCTTGCTCAG TCAGATGACC	2460
TATACCCGTG TTCCCGTTGT GACAGATGAA AAACAGTTT TGGGACGAT TGGACTCAGA	2520
GATATTATGG CTTATCAGAT GGAGCATGAC TTGAGCCAAG AAATCATGGC GGATACGGAT	2580
ATCGTTCATA TGACAAAAC GGACGTAGCG GTTGTTTCGC CTGATTTCAC CATTACGGAG	2640
GTCTTGACA AGCTAGTAGA TGAGTCCTTC TTACCGGTTG TGGATGCAGA GGGTATTTTC	2700
CAAGGGATTA TTACGCGCAA GTCCATCCTC AAGGCCGTTA ATGCCCTCTT GCATGACTTT	2760
AGTAAGGAAT ATGAGATTCG ATGCCAATGA GAGACAGGAT TTCAGCCTTT TTAGAGGAAA	2820
AGCAGGGCTT GTCTGTCAAT TCCAAGCAGT CCTATAAGTA TGATTGGAG CAATTTTTAG	2880
ACATGGTAGG TGAGCGGATT TCTGAGACCA GTCTCAAGAT TTACCAAGCC CAGCTAGCCA	2940
ATCTAAAAAT CAGCGCCAG AAGCGAAAGA TTTCCGCCCTG TAACCAATTT CTATACTTTC	3000
TCTATCAAAA AGGAGAGGTG GACAGCTTTT ACCGCTTGGA ATTAGCCAAA CAAGCTGAAA	3060
AGAAGACGGA AAAGCCAGAG ATTCTATACC TAGACTCTTT TTGGCAGGAA AGCGACCATC	3120
CAGAGGGCCG CTTGCTAGCG CTCTTAATCC TAGAAATGGG GCTCTTGCCC AGTGAGATTT	3180
TAGCCATCAA GGTGCGGAC ATCAATCTGG ATTTTCAGGT GTTGCGAATC AGCAAGGCTT	3240
CCCAACAGAG GATTGTCACC ATTCCCACGG CCTTGCTTTC AGAATTGGAA CCCTTGATGG	3300
GGCAGACCTA TCTTTTTGAA AGAGGAGAGA AACCTATTTC TCGTCAGTGG GCCTTTCGTC	3360
AGTTAGAATC TTTTGTCAAG GAGAAAGGTT TTCCATCCTT ATCAGCTCAA GTCTTACGTG	3420
AACAGTTTAT TCTAAGACAA ATAGAAAACA AGGTCGATTT GTACGAAATT GCAAAAAAAT	3480
TAGGATTAAA AACAGTCCTG ACCTTAGAAA AATATAGATA ATGGATATTA AATTAAAAGA	3540
TTTTGAAGGA CCCCTGGACT TGCTCTTGCA TCTGGTTTCT AAGTACCAGA TGGATATCTA	3600

1060

CGATGTGCCC ATTACGGAAG TCATCGAACA GTATCTAGCC TATGTCTCAA CCCTGCAGGC	3660
CATCGCTCTG GAAGTGACGG GTGAGTACAT GGTCTGGCT AGTCAGCTCA TGCTGATTAA	3720
GAGTCGTAAA CTCCTTCCGA AGGTAGCAGA AGTGACAGAC TTGGGGGATG ACCTGGAGCA	3780
GGACCTCCTC TCTCAAATCG AAGAATATCG CAAGTTCAAG CTCTTGGGTG AGCACTTGGA	3840
AGCCAAGCAC CAAGAACGGG CCCAGTATTA TTCCAAAGCG CCGACAGAGT TGATTTACGA	3900
AGATGCGGAG CTTGTGCATG ACAAGACGAC CATTGACCTC TTTTGTGACTT TTTCAAATAT	3960
CCTAGCCAAG AAAAAAGAGG AGTTTGACA AAATCACACG ACGATCTTGC GGGATGAGTA	4020
TAAGATTGAG GACATGATGA TTATCGTGAA AGAGTCCTTG ATTGGACGAG ATCAATTGCC	4080
CTTGCAGGAT TTGTTCAAGG AAGCCCAGAA TGTCCAAGAG GTCATCACCC TCTTTTGGC	4140
AACCCTAGAG TTAATCAAAA CCCAGGAGTT GATCCTCGTG CAAGAGGAGA GTTTTGAGA	4200
TATCTATCTC ATGGAAAAGA AGGAAGAAAG TCAAGTGCCT CAAAGCTAGA CTTGATAGAG	4260
AGGAAAGATG AGTACTTTAG CAAAAATAGA AGCGCTCTTG TTTGTAGCGG GTGAAGATGG	4320
GATTCGGGTC CGCCAGTTAG CTGAACTCCT CTCTCTGCCA CCGACAGGCA TCCAGCAAAG	4380
TTTAGGAAAA TTAGCCCAGA AGTATGAAAA GGACCCAGAT TCCAGTTTGG CTTTGATTGA	4440
GACAAGTGGT GCTTATAGAT TGGTGACCAA GCCTCAATTT GCAGAGATTT TGAAGGAATA	4500
CTCTAAGGCG CCTATCAACC AGAGCTTGTC TCGGGCTGCC CTTGAGACCT TGTCCATTAT	4560
TGCCTACAAA CAGCCGATTA CGCGGATAGA AATTGATGCC ATCCGTGGAG TTAACGCGAG	4620
TGGAGCCTTG GCAAAGTTGC AGGCTTTTGA CCTGATAAAG GAAGACGGGA AAAAGGAAGT	4680
ATTGGGGCGC CCCAACCTCT ATGTGACTAC GGATTATTTT CTAGATTACA TGGGGATAAA	4740
CCATTTAGAA GAATTACCAG TGATTGATGA GCTTGAGATT CAAGCCCAAG AAAGCCAATT	4800
ATTTGGTGAA AGGATAGAAG AAGATGAGAA TCAATAAGTA TATTGCCCAC GCAGGTGTGG	4860
CCAGTAGGAG AAAAGCAGAA GAGCTGATTA AGCAAGGCTT GGTGACGGTT AACGGCCAAG	4920
TGGTGCGTGA ACTAGCAACC ACTATCAAGT CAGGCGACAA GGTGGAAGTT GAAGGTCAAC	4980
CTATCTACAA CGAAGAAAAG GTCTACTATC TGCTTAACAA ACCACGCGGT GTGATTCCA	5040
GTGTGACAGA TGATAAGGGT CGCAAGACGG TTGTCGACCT CTTGCCCAAT GTCAAAGAGC	5100
GTATTTACCC TGTGGGTCGT TTGGACTGGG ATACATCAGG TGTCTTGATT TTGACCAATG	5160
ATGGGGACTT TACAGACGAG ATGATTCACC CTCGTAATGA GATTGACAAG GTTTATGTCTG	5220
CGCGTGTTAA AGGTGTGGCC AATAAGGACA ATCTCCGCCC CTTGACCCGT GGTCTTGAGA	5280
TTGATGGTAA GAAAACCAAG CCAGCTGTTT ATGAAATTCT CAAAGTGGAC CCAGTCAAAA	5340
ATCGCTCTGT GGTGCAGTTG ACCATCCATG AAGGGCGTAA CCATCAGGTT AAAAAGATGT	5400

1061

TTGAAGCTGT TGGTCTCCAA GTAGATAAGT TGTCTCGGAC TCGTTTCGGA CACCTAGACT	5460
TGACAGGACT CCGTCCAGGA GAATCCCGTC GTCTTAATAA AAAAGAAATC AGCCAACCTAC	5520
ACACCATGGC TGTAACCTAAG AAATAATGAA ACGAATTTTA ATAGCGCCTG TGCCTTTTAA	5580
CCAACGTTTT ATCTCACCAG TCTTTCCACC CTCTTGTCGC TTTGAGCTGA CTGTCTCCAA	5640
CTACATGATT CAGGCTATTG AAAACATGG GTTTAAGGGG GTATTGATGG GCTTGGCTCG	5700
GATTTTACGT TGTATCCCT GGTGAAAAC AGGTAAGGAC CCGTTCCAG ACCGCTTTTC	5760
CCTTAAACGA AATCAAGAAG GGAATGAGG TGGGGTAAAT AGATTTCAAA ATGATAAAAA	5820
CGCATCCTAT CAGGTTTGAG TGAACCTGAT AGGATGCGTT TTAGAATGTC AAAATTTTAT	5880
ACTCTTCGAA AATCTCTTCA AACCCTGTC GCTTTCATCT GCAACCTCAA AACAGTGTTC	5940
TGAGCAACCT CCGGCTAGTT TCCTAGTTTG CTCTTTGATT TTCATTGAGT ATTAATTTGA	6000
GTTTGAAGTG GCTTATTTCA AAGCTTTTTG TATGTCTTCA ATCATGAGTT TTGTGATTC	6060
AAGTCCCGCT CCGCTTAGAT ACCAGAGGTC TGGTGTAGT TGGATAATCT TACCATTTT	6120
AGCAGCAGGT GTTTCAGCGA TAAGGGCATT TTCTAGGACA CCGTCCTTGC TAGAGTTGTC	6180
CCCACCGATG GCAAGGTAC GGTGATGAC AAAGAGGATG TCAGGGTTGA TTTCTTTGAC	6240
ACTTTCAAAG CTGACTTCTT GTCCGTGGCG TGAGTCTTCA AATTTGTAT CAGTTGGTTT	6300
GAATTTCAAG GTTTGGTACA AGAAAGAGAA ACGAGATTG GCACCAAAG CTGCCATTTT	6360
TCCTTCATTA AGGAGGATCG CAAGGGCTTT TTTGTGAGAG CTTTCATTTT TAGTAGCGAC	6420
TTCTTGGATG CTCTGTCTA GCTTGGTCAA TTCTTCCTTG GCTTCTGTG TACCAGTTTC	6480
GCCGAAGGCA CTTGCTAAGG ATTCGATATT AGCCTTGGTA GAAGTCCAGT AGTCGTCCTT	6540
GCTTGCTTGG AAGAGAACGG TTGGGGCGAT TTCTTTGAAT TTGTCTACGA ATTTTGTGT	6600
ACGTGGCGAA GCGATAATCA AATCAGGCTC AAGGGCGGCG ATAGCTTCTA AATCAGGTTC	6660
TTTCATAGAA CCAACATTTT TGACAGTTCC CACTAGGTCT TTTAGATAAG TCGGAACAGT	6720
TTTGTAGGC ATTCCGACGA TATTTTTTTC AAATCCTAAA GCGCGAATAG TATCCGCAGC	6780
GCCGAGGTCA AAGGTCACAA TCTTTTCAGG AACTTTGGAA AGTTTGACCT COTCCAGTGA	6840
ACTTTTAATG GTTACCTCTG TTGGAGCAGA GCTACTGGTC TCTGTCTGAC TAGTGCTTGA	6900
GTGTGTACTA CATGCACCAA GTAGGAGCAA GAAGCTGGCC ACTAGGGCAG TGAAATAAAG	6960
TTTAAGGGAT GTTTTCATAA TTTCTCCTTT TTTAAATGTG ATAACGATTT AGGGAGTCTC	7020
TTAATCTTAT TGAATAAGAG ACTGAAGTT CTCTAACTTG ACCTTTTATG TTAGTAGCTA	7080
TAGATACAGA TCTTTTGTG ATTGATATCA GCTAGCGTGA TGGGAATCTC ATAAAGTTGA	7140

1062

CTGAGCAGGT CAGCCTGCAT GATTTGATCG GTTCTTCCCT TGCTAAAGAC CTGGCCGTCC	7200
TTGAAGGCGA CAATTTTCATC TGCATACTGA CTGGCCATGT TGATATCGTG GAGGACGATG	7260
ATAATGGTCT TGCCGAGTTC CTCCACCAGT CGTCGAAGAA TCTGCATCAT GCTGACGCTT	7320
TGCTTGATAT CGAGATTGTT GAGTGGTTCG TCCAGCAAGA TAAAGTCCGT ATCCTGGGCC	7380
AGTACCATAG CGATAAAGAC GCGCTGGAGT TGCCCCCCTG ACAGGCTATT GATGTAGCGG	7440
TCTTTTAAGT TGGTCAGTTC TAAATAGTTC AGAGTTTCTC GGATTTTTTC CCAGTCTTCT	7500
GATCTAAGTC GACCTCGGCT GTAGGGAAAA CGTCCAAAAC TGACCAGTTC TTCAACAGTC	7560
AATTTGGCTT GGTAATTGAT TTTCTGTTTT AGGATGGTTA GTTCTTGGGC CAGTTCTTGC	7620
GAATTCACG TCTCGATTC ACGTCCTTTG ATACTGAGAA CTCCCTGATC TTTCTTGGTT	7680
AGCCTGCTCA TGATGGAGAG GAGAGTCGAT TTTCCAGCAC CATTTGGACC AATAAAGGCT	7740
GTCAGTTTTT GAGGACTGAC TTCAAGCGAA ATGCCTTGCA AAATATCCTG TTTTGAATG	7800
GATTTGTCAA TGTTTTCCAG TTCACTGAC GAGACCTCCT ATATAGTAAG ATAAAGAATA	7860
AGAAGCCACC CACACTCTCA ATGATCATAC TGATACGAAT TTCCAGTGCA AAGACTCGTT	7920
CAATCAAGGC TTGCCCCAAG GTTAAGCTAA TAAATCCAAC CAGAATGGCC ACTATAAAGA	7980
GTAACCTGTG CTGATAGTCT TTGACAATCA GGTAGGTGAG GTTGGCCAGT ATAAAGCCGA	8040
AGAAGGCCAT AGGTCTTACC AAGGCAGTGG CCGTTGAGGT CAAAAGCAGC ATTCCCCAGA	8100
GGAGCTCTTT CTGTTCTTTT TCAACATCGA GTCCCAATAT CTGAGCCGTT TCTCTTTGCA	8160
GGTGCAAGAC ATCTAGAACG ACTGCTTTTC GAAAGAAAAA GATTGTCAA GCGAGGATGA	8220
TCAGAGAACC GATGGCTAGG ATGGAAGTGT TGAGATGTTG AAAGGAGGCA AAAAGACTAT	8280
TTTGCAGTTT ATCGTATTCG TTTGGATCCA TTAGGACTTG AAGGAAGGTG CTGATATTTT	8340
GAAAGAGACT TCTGAGCGCT AGACAGATCA GCAGGACGAA GACCAGGTCT TGCTTCATCA	8400
GTGTCTTCAA GTAACCTTGT AAGGCGAGAA AGAAGAGGGA CTGGACAAGA AGTAAGACTA	8460
GGAATTCTAA GATAGGGGAT TTGCCAAGTT GAAGAACTT GCTTTCAAAA ACCAGTAGTA	8520
GGGTTTGTAG TAGGACGTAG AAGGATTCAA TTCCCAAAAT ACTAGGCGTC AGGAAGCGAT	3580
TTTCCGTCAG GGTTTGAAAA CTAATGGTCG AAATCCCAGT CGCGATGGCT ACCAAGAGAT	8640
AAACGATGAT CTTTGGGAA CGCAACTTCC AAGCAAAGGC TGACAAGTGA GTGATGGGCC	8700
AAAAGTAGAG AAGACAAGCT CCGATGGCAA GAATAATGAG AATCCAGAAG AGCTTGGTAT	8760
GTTTGCTTTT AGTCTGCATC TTTTCGTCCC CCTCTCCAGA GAAGTAGGAT AAAGACGAGA	8820
CTACCGATGA TTCCTAGCAA GAGACTGACA GACAACTCAT AGGGCCTAAT CAGAACTCGG	8880
GATAGGATAT CGCAAGCCAG AACTAGATTG GCACCAACCA GTGCGACCAT GAGTTTGGTT	8940

1063

TGACTTAGAT TATCTCCATA GCGCTTGCGA ACAAGATTGG GAACGATAAC TCCGAGAAAT	9000
GGTAGGCCAC CCACGGTAAT CATGGTGACG CTTGTCTGTA GCGCCACCAG AAAGAGGGCC	9060
AGTTTTTCAA GTAGGGAGTA GGAAATCCCC AAACCTCTCG TGGTTTCTTT CCCTAGATTC	9120
ATGATGGTGA AGGTTTGGA TAATTTCCAA ACGGTTATCA GGATGATGAG GCCTAAGAAG	9180
AGCCACTCAT ACTGATGGGT CTGAATCATG GAGAAGGAGC CCTGGGTCCA GGCAGTCATA	9240
CTCTGAACCA GATTGAAACG ATAGGCGATA ACTTCTGTGA CTGAGCCGAT AATCCCGCTA	9300
TAGATGATCC CAATCAGAGG CAACATCCAC CTTTCCTTTA CAGTAAAAAT GGTCATAAAG	9360
GCTAGGAAGA AGAGGGTGAA TACGATGGAT GAAACAAAAG CGAAGAGCAT CTTGTGGGTC	9420
AGACTAGCCG ATGGAAAGAC AAAAAGGCTC AGCACCATTG CCAGTTTGGC GGCTTCAGTC	9480
GTTCCAACTG TACTCGGTGC AGCAACTGA TTTTGGGTAA TAGTCTGCAT GAGAAGGCCT	9540
GCCATACTCA TACTAGAGGC AGTCAGGAGA ATACTGATAG TTCTTGGGAG ACGGGACTCT	9600
TGAAAGAGGA GCCAGGTCTG CTGGTCGAAA TCAAATAGCT TTCCCATGA AAAATCACTG	9660
GTCCCAATGC TAATAGAGAG AAAGACTAGG AGTAGAAGTA AGCCAGG	9707

(2) INFORMATION FOR SEQ ID NO: 165:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5910 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

CCGCAATTAT GCTTGAAAAG GAGTATACTT ATAAGTAACG CAAACGTTTG CGTCTGAAAA	60
ATACGCAACG TTCCATTATT TTAACACACG AGGTGCTATT ATGAAAAAC GTCAAAGTGG	120
TGTGTTGATG CACATCTCTT CTCTTCCAGG AGCTTACGGA ATCGGATCAT TTGGTCAAAG	180
TGCTTACGAC TTCGTTGATT TCTTGGTCCG TACAAAACAA CGTTACTGGC AAATCCTTCC	240
ATTAGGAGCA ACTAGTTACG GGGATTCTCC TTACCAATCT TTCTCAGCCT TCGCAGGAAA	300
CACTCATTTT ATCGATTTAG ATATCTTGGT GGAGCAAGGT TTGTTGGAAG CAAGTGACCT	360
TGAAGGAGTT GACTTTGGTA GCGATGCGTC TGAAGTTGAC TATGCTAAAA TCTACTATGC	420
ACGTCGTCCT CTTTTAGAAA AAGCGGTGAA ACGTTTCTTT GAAGTCGGAG ATGTTAAAGA	480
TTTGTAGAAA TTTGCTCAAG ACAACCAATC ATGGCTTGAG CTCTTTGCTG AGTATATGGC	540
TATCAAAGAG TATTTTGACA ATCTTGCTTG GACTGAATGG CCAGATGCAG ATGCTCGTGC	600

1064

TCGTAAAGCT TCAGCACTTG AAAGCTATCG TGAGCAATTG GCAGACAAGT TGGTTTACCA	660
CCGTGTGACT CAATACTTCT TCTTCCAACA ATGGTTGAAA TTGAAAGCTT ACGCTAACGA	720
CAACCACATC GAAATCGTTG GGGACATGCC AATCTACGTA GCGGAAGATT CAAGTGATAT	780
GTGGGCAAAT CCACATCTCT TCAAAACAGA TGTCAATGGT AAGGCTACTT GTATCGCAGG	840
ATGCCCACCA GATGAGTTT CTGTAAGTGG TCAGCTTTGG GGTAATCCAA TCTATGACTG	900
GGAAGCAATG GACAAAGACG GCTACAAATG GTGGATTGAA CGCTTGCGTG AAAGCTTCAA	960
AATCTACGAT ATCGTTCGTA TCGACCACTT CCGTGGCTTC GAATCTTACT GGGAAATCCC	1020
TGCTGGTTCC GATACAGCAG CACCTGGTGA GTGGGTGAAA GGTCCAGGTT ACAAGCTTTT	1080
TGCAGCCGTT AAGGAAGAAC TTGGTGAGCT AAACATCATC GCAGAAGACC TTGGCTTCAT	1140
GACAGATGAA GTGATCGAAT TGCCTGAACG TACTGGCTTC CCAGGAATGA AGATTCTTCA	1200
ATTTGCCTTC AACCCAGAAG ACGAAAGCAT TGATAGCCCA CACTTGGCAC CTGCTAACTC	1260
AGTTATGTAC ACAGGAACAC ACGATAACAA TACGGTTCTT GGTTGGTACC GTAATGAGAT	1320
TGATGATGCG ACTCGTGAGT ACATGGCTCG TTACACGAAC CGTAAAGAAT ACGAAACAGT	1380
GGTACACGCT ATGCTTCGTA CAGTATTTTC ATCAGTTAGC TTTATGGCAA TTGCAACTAT	1440
GCAAGATTTA CTAGAATTGG ATGAGGCAGC TCGTATGAAC TTCCCATCTA CCCTTGGTGG	1500
AAACTGGTCT TGGCGTATGA CTGAAGATCA ATTGACACCA GCTGTCGAGG AAGGTTTGCT	1560
TGACTTGACA ACAATTTATC GCCGAATTAA TGAAAATTTG GTAGATTTAA AGAAATAAGA	1620
CAATAATCAG GAGACAACTA AACATGTTAT CACTACAAGA ATTTGTACAA AATCGTTACA	1680
ATAAAACCAT TGCAGAATGT AGCAATGAAG AGCTTTACCT TGCTCTTCTT AACTACAGCA	1740
AGCTTGCAAG CAGCCAAAAA CCAGTCAACA CTGGTAAGAA AAAAGTTTAC TACATCTCAG	1800
CTGAGTTCTT GATTGGTAAA CTCTTGTCAA ACAACTTGAT TAACCTTGGT CTTTACGACG	1860
ATGTTAAAAA AGAACTTGCA GCTGCAGGTA AAGACTTGAT CGAAGTTGAA GAAGTTGAAT	1920
TGGAACCATC TCTTGGTAAT GGTGGTTTGG GACGTTTGGC TGCCTGCTTT ATCGACTCAA	1980
TTGCTACTCT TGGTTTGAAT GGTGACGGTG TTGGTCTTAA CTACCACTTT GGTCTTTTCC	2040
AACAAGTTCT TAAAAACAAC CAACAAGAAA CAATTCCAAA TGCATGGTTG ACAGAGCAAA	2100
ACTGGTTGGT TCGCTCAAGC CGTAGCTACC AAGTACCATT TGCAGACTTT ACTTTGACAT	2160
CAACTCTTTA CGATATTGAT GTTACTGGTT ATGAAACAGC GACTAAAAAC CGCTTGCGTT	2220
TGTTTGACTT GGATTCAGTT GATTCTTCTA TTATTAAAGA TGGTATCAAC TTTGACAAGA	2280
CAGATATCGC TCGCAACTTA ACTCTCTTCC TTTACCCAGA TGATAGTGAC CGTCAAGGTG	2340
AATTGCTCCG TATCTTCCAA CAATACTTCA TGGTTTCAAA CCGTGCGCAA TTGATCATCG	2400



1065

ACGAAGCAAT CGAAAAAGGA AGCAACTTGC ATGACCTTGC TGACTACGCA GTTGTCCTAAA	2460
TCAACGATAC TCACCCATCA ATGGTGATTC CTGAATTGAT TCGTCTTTTG ACTGCACGTG	2520
GTATCGATCT TGACGAAGCA ATCTCAATTG TTCGTAGCAT GACTGCCTAC ACTAACCACA	2580
CAATCCTTGC TGAAGCGCTT GAAAAATGGC CTCTTGAATT CTGCAAGAA GTGGTTCCTC	2640
ACTTGGTACC AATCATCGAA GAATTGGACC GTCGTGTGAA GGCAGAGTAC AAAGATCCAG	2700
CTGTTCAAAT CATCGATGAG AGCGGACGTG TTCACATGGC TCACATGGAT ATCCACTACC	2760
GATACAGTGT TAACGGGGTT GCAGCACTCC ATACTGAAAT CTTGAAAAAT TCTGAGTTGA	2820
AAGCCTTCTA CGACCTTTAC CCAGAAAAGT TCAACAACAA AACAAACGGT ATCACTTTCC	2880
GTCGTTGGCT TATGCATGCT AACCCAAAGT TGTCTCACTA CTGGATGAG ATTCTTGGAG	2940
ATGGTTGGCA CCATGAAGCA GATGAGCTTG AAAAAGTTTT GTCTTATGAA GACAAAGCAG	3000
TTGTCAAAGA AAAATTGGAA AGCATCAAGG CTCACAACAA ACGTAAATTG GCTCGTCACT	3060
TGAAAGAACA CCAAGGTGTG GAAATCAATC CAAATTCTAT CTTTGATATC CAAATCAAAC	3120
GTCTTCACGA GTACAAACGC CAACAAATGA ACGCTTTGTA CGTGATCCAC AAATACCTTG	3180
ACATCAAAGC TGGTAACATC CCTGCTCGTC CAATCACAAT CTTCTTTGGT GGTAAAGCAG	3240
CTCCAGCCTA CACAATCGCT CAAGACATTA TCCATTTAAT CTTTGCATG TCAGAAGTTA	3300
TTGCTAACGA TCCAGCAGTA GCTCCACACT TGCAAGTAGT TATGGTTGAA AACTACAACG	3360
TTACTGCAGC AAGTTTCCTT ATCCCAGCAT GTGATATCTC AGAACAAATC TCACTTGCTT	3420
CTAAAGAAGC TTCAGGTAAT GGTAAATGA AATTCATGTT GAACGGAGCT TTGACACTTG	3480
GTAATATGGA CGGTGCTAAC GTGGAAATCG CTGAGTTGGT TGGAGAAGAA AACATCTACA	3540
TCTTCGGTGA AGATTCAGAA ACTGTTATCG ACCTTTACGC AAAAGCAGCT TACAAATCAA	3600
GCGAATTCTA CGCTCGTGAA GCTATCAAAC CATTGGTTGA CTTATCGTT AGTGATGCAG	3660
TTCTTGCAGC TGGAAACAAA GAGCGCTTGG AACGTTTTTA CAATGAATTG ATCAACAAAG	3720
ACTGGTTCAT GACTCTTCTT GATTTGGAAG ACTACATCAA AGTCAAAGAG CAAATGCTTG	3780
CTGACTACGA AGACCGTGAC GCATGGTTGG ATAAAGTCAT CGTTAACATT TCTAAAGCAG	3840
GATTCTTCTC ATCTGACCGT ACAATCGCTC AGTATAACGA AGACATCTGG CACTTGAAC	3900
AATACTCTTC GAAAATCTCT TCAAACCACG TCAGCTTTAT CTGCAACCTC AAAGCAGTGC	3960
TTTGAGCAAC TGCGGCTAGC TTCCTAGTTT GCTCTTTGAT TTTCAATTGAG TATAAGATAC	4020
AAATTTATAC TAATACATTT TGTAAAAAAG CGAGTTTCGA TTGAAATTCG CTTTPTTAAT	4080
GATGTAGATT TGGGTCAATC TTGTCTAAAA ATAGGGAAAT CCTAGATACA GTGAAGGCTT	4140

1066

TAAATGCTGG TTTTACTGT CCTCAGCCTT ATATTTTTTC GTAGTTGGTT ACCTCATATC	4200
TATTATATTC GCTTACATAA AGTATTATAA TATAATTGTA GGAAAGAAGG TGTTTTTATG	4260
ATATACACAC TTAAATTGGT GTTGTATTAT ACCTTTCTTG TAATAAGCTT GTTACCTGAT	4320
AAGATTTTTG GAAAAATAA AAAAATTGG AAAATAGTTT TTGCAATATT GACGGCAGTG	4380
GCAGCATTGT CATTTATGTA CTAAGTTATT TTAAGAATGT AGGGAAATAA ACCCTACATT	4440
CTTTTTAGTT TTTTCTGTTT TCTAAATTCT ATTTATCCAA GCGATTCAAC ATTTCTTGCT	4500
TCTTCGCTTC AAGTCTGCA CGCTTTTCTT CGATTTCCGC ATGTTTTTTC TCGAGTTCAG	4560
AACAACCTGC ACCATTGCTA AATTCTTTTC GCCATCAGGA GATAGGGTGA GTCGACATGT	4620
CTATTACTCA CCCAAAGCAG TCCTACAAAG CAGGAATTTT CTGTTACTTT TTTGGAAATA	4680
GTAACGTTTA TACAGCTTTG AACTTTCGTA TCAAAGCGCC AAACACACTC CGAGGGGTTT	4740
ACAGAAAGCA GAAAAGGAAT GATCTGGTAT AAGATCATTC CTTTTCTCTC TTTTTCTTTA	4800
AGTAATTATA TACAATGTAC GACGAAGTCG TCATTGCAAT GCTGATCCAC CACCTAAAGG	4860
GAACTTTAAA CAACATTGAT AAGATAAAGA ATATAACAA CGAAAATACG TTATACCCAA	4920
TTAATTTTAT TGTATATCTC ATGATTAATA GTTAATCCTT CCGTTGTTAG GAATGGCATC	4980
ATTTTTATCC CATAATTGTG CTAAATAAGT CCCCAGTGAT AATAAATCA TAGCGAATTC	5040
TAAAGCAACA TCATTTACAA ACCAACTACC TAGATATCTA GAAATTGCTG AACGAATAGC	5100
ACTTTTTGCT GCATGTTTC CTTTTACTTT AATTAGATTT GCAAGGCCTG CAGTAGTTCC	5160
TCCTAATGCT AAAGCTATG CAGTATCTAA TAGAGCACCC ATTTGATTAA CTGTAATACC	5220
TTGCCAACT GCTCTAAATG GAGAGTATGT AGGTGGGATT GTATAATCGC CTTGTAATTG	5280
TCGGTTAATT ACTTCTTTGA TCCATTGTTG TGAGACGTCT GGATGAAAAG ATTGGATTTT	5340
GTTTGCAAGT GTATTGATTT GTTCTTCTGT TAGAGAAGTG ACAGGTTGAA GTTCCATATT	5400
TGTTTCAATT TGTGATACTT GTTCAGAAGC GTATACAGCT GAAACACTTG GAATCGCTGA	5460
TACAATTAAC ACAATTGACG TCAAAAAAAC CGAAATAAAT TTCATTAATT TGTTCATGAG	5520
CTTTTCTCCT TTTTATTTGC ATCTGCTTAC ATTTTATCAT ATACTGTTAT TATAGTCAAA	5580
AAAATATGCT ATTATGTTAA AAAAATATTT TTCAAATAT AAATGGACGG ATTTATTTTG	5640
GATTTTATTT GTTATTTTGA CCTGCCTCTA TATTGGTAAC CATGATTGT TTAATCTCAA	5700
TCATCAAGAA TTCTCTTTTC GTGGTAGCGT TTGGGGTCTG GTACTGGCCT TATATCACTT	5760
ACTATTCATT GATAAGTTTG TTATATCGAA TCGAAAATAA AGATTAGAGC TATGCTTGAC	5820
TGTGTACTTT TAGGATTTAT TTTGGAGGAA GATTTTGTCT CTATTATTTA TTATTTTAAA	5880
TTTATTTATT TTGTATAAGA TCTATTCTTT	5910

1067

## (2) INFORMATION FOR SEQ ID NO: 166:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5406 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

GGCATAGCGA CTCATTTTTT CAACTGTCCA GGCTGGATAC CAGACTAATT TAACCTCAGT	60
ATCCGTTACT TCTGGAACCT CTATCATAGC ATCATAAATC TGGTCTGTCA AAAGGTCTGC	120
TAAGGGACAA CCCATAGTTG TCAAAGTCAT GTCAATCTCT GTTTGCCCTG TGTACCCTC	180
AAAACGAATC TCATAGATCA AACCAAGATT GACAAATATCG ATTCCCAACT CAGGGTCGAT	240
GACTTCTTCC AAGGCTGTTA AAATCCGTGT TTTGATGTTT TCAATTTGCT CTTCTGTATA	300
AGCCATATTT TCCTCACTCT TAGTCTTCAA TAAATCAGG AAGCGGTTTG CTACGACTTG	360
GTTGGCGTAG TTTTCTCAAA GCCTTTGCTT CAATCTGACG GATACGCTCA CGAGTTACGT	420
TAAAGACTTT CCCCACATCT TCAAGTGTGC GCATTTTCC ATCATCTAGT CCAAAACGTA	480
GACCGAGAAC ATTTTCTTCA CGGTCTGTAA GAGTATCTAA GATTTTCATCC AATTGCTCAC	540
GCAAGACGAT ACGAGTCGTA TAATCCACTG GATTTTCAAT CACTTCATCT TCGATAAAGT	600
CTCCAAGGTG GCTATCGTCC TCTTCACCGA TAGGAGTTTC AAGAGATACT GGTCTTGGG	660
CAATCTTCAA GATTTTCACGA ACCTTATCAG GTGTCATATC CATTTCGTTCA GCAATCTGTT	720
CTGGTGTCCG ATCTTGCCCC AATTCTTGAA GGAGATTCCG CTGTTTCACGA ACCAATTTAT	780
TGATAGTTTC AACCATGTGA ACTGGGATAC GGATGGTACG AGCTTGCTCC GCAATAGCAC	840
GAGTGATAGC CTGACGAATC CACCAAGTTG CATAAGTTGA AAACCTGAAC CCTTTAGAAT	900
AGTCAAACCT GTCAACCGCC TTCATCAAGC CCATATTTCC TTCTTGAATC AAGTCAAGGA	960
ACTGCATACC ACGACCGACA TAGCGTTTGG CAATGGAAAC AACCAAACGA AGATTGGCTT	1020
CCGCAAGACG TTGTTTGGCT TCGATATCAC CAGCTTCAAC AGCCAGTGCC AACTCTTCT	1080
CCTCTTCATT GGTCAAGAGA GGAACGACCC CTATTTCTTT CAAGTACATA CGGACAGGGT	1140
CATTGACCTT AGCAGAAGTT GACCCAATCA AGTCCTCATC GCTGAGTTCT GGTCTTCTT	1200
CATTGCTGAG AACACGCGCA CTTGGATTTC CTTTCGTTATC TGTGATAGAA ATGCCTGCAT	1260
CCTGAATCCG TTGCAAGAGA TCTTCAATCC CATCAGCGTC CAAGGTAAAA GGAATAACCA	1320
GACTTGCATT GATTTTCATCA TCTGTTGCTG TCCCTTTTGT CTTATGATTA CGGATAAATT	1380

1068

CTGCTACCTG TACGTCAAAT GTTGTTACTT CTTTTTGTTT TGTGCCCATT ATTACTCCAT	1440
TCTTCTCTTT TGGGAAATTA AACGTTCCAA TTCTTCTAGG GCTGTATCTG TATCTCCTAC	1500
ATGGCTAGCT TCCTGCACCT TCTTTTTGAT TCTCATATTG TCCTGATTCA AGAGAGCCTT	1560
GTTTCGAGTC ATCTCTACTT CACTAAGTTC CTGCGCGCAT ATCTCAGCAG GCAAATCCTG	1620
AGCTAAAACT TGGTACCAAG CTCTTTCAAC TTCCTCTGTC TGCTCTGCTA AAACCTCTGG	1680
AGGAAGATTT CCATACTGGC CAAGCAAGTC ATATAAGACC TGAAATTCAG GTGTAGCAAA	1740
TGCAAAGTCT TCTCGCAAAC GGTAAATCGTT CAAAACAAGA GGGGATTCCA TCATCCGATA	1800
GAGTAGATGG GCTTCTGCCC TCATAATAGC CGATAACTGC TTGGTGACAG GCATGGTGAT	1860
TGGCGTCGGT CTGGAAATTC CTTCCATGCG ATTCTGCCTT TGCACCTGAC GACTCTCATT	1920
AACAATCTGC TCAATCTGGG TATAATCAAA GGACGCCAGA CTGTCAGCTA AAATATGAAT	1980
ATAGCTGTTT TGAGCAGCGA TGGACTTTTC TTGAACAATC AAGGGAGCTA TTTTTTCAAG	2040
AAACTCAATC TGAGCCTGCA GATTTTCACT GTTTTCAGGT TTGTACTGAT GAATGTAGAA	2100
CTCAATCGGA CTAATACGAG TTTTCGTTAA TAGATAGGCC AAGTCTTCTG GACCATTTTT	2160
TTGTAGATAC TCATCAGGAT CCAAGTTATC AGGCATGCTG ACGATTTGCA CAGGCATATC	2220
ACCAATTTCA TCCAATGCTT TCAATGTCGC GGCTTGCCCA GCCTTATCTC CATCGTAAAC	2280
AAGAACCAAT TTCTTGGTTA ACCTTTTCAG ATGCTCAACA TGCTCTCGAC TCAAGGCTGT	2340
TCCCATCGAC GCCACAGCAT TTTGATTCC AGCCCGATAG GCTGCAATAA CATCCATGAA	2400
TCCTTCCATC AGGTAAATCT CACTAGCTTT TCCAGAAGAT CTTTTTGCCC TATCCATATG	2460
ATATAATTCG TAACTTTTGT TAAAAATTGC AGTCGATCGG CTGTTTTTAT ACTTAGAAGT	2520
TTGTGAATCC GTTTTTTGCC AGATACGACC TGAGAAGGCA ATGACCTTTC CTTGGTCATT	2580
TGTCAGGGGA AACATAATGC GATTGTGAAA GGTGTCTACA AATTGATTGG CATCCGAGAG	2640
ATAAACAGG CCTGAATCCA GTAAATCCTC TTCACGATAC TGATCAGACA AACGTTGATA	2700
GAGATAGTTT CGTTCTGGAG GTGCTAAACC AATCCAAAAA TGTTTAAGCA CTTCATCTGT	2760
CAACCCCCGC TGATAAAGGT AATTTCTGGC CTCTTCGCCC ATAGTCGTTG TCATGAGAAT	2820
AGCATGGTAA AATTTGGCTG CATCTTCGTG CATATCATAA AGAGCTTGGT GAGGTGAGGC	2880
TGACTTCTGC TCACTATAAA GCGGTTTTTC AACCTCAATT CCAACACGCT GACCTAAGAT	2940
TTGGACTGCT TCTATAAAGG GAACCCCTTG GTACTCCTCG ATGAACTTAA AGACATCACC	3000
TGAGCGACCA CAACCGAAAC AGTGATAAAA CTGCTTGTC TCTACAACAT TGAAAGATGG	3060
TGTTTTTTCA CCATGAAAAG GACAGAGCCC TAGATAGTTC CGTCCTGCCT TTTGTAAAGA	3120
AATCACATCT CCTATGACTT CCACAATGTT GGCATTGTTT TTGATTTCTT CAATGACTTG	3180

1069

TTTGTCAACC ATACACAATA CCTCCATGTT ATCATAGTTT ACTTTATATA GTATACTTTA	3240
TTTCAGAAAA AAAGTAAACC ATTTCACTCA TTTTCCCTAC TTTATTCAAA GAGTTGATAA	3300
TAATCAGAGA TTTTCATTTT TGCTTTTCT TCTTGGTTTA AATCTTGGAT AATTCGTCCT	3360
TCTTTCATGA CAATCAAGCG ATTGCCGTAT TTGAGAGCAT CTTCCATATG ATGAGTAATC	3420
ATAAGGGCTG TTAGCTGATC TTTCTTAACA AATTCATCTG TCAATTCCAT CAAAGCAACA	3480
CTAGTCTTTG GATCCAGGGC AGCAGTATGC TCATCTAACA GGAGTAATTC AGGTCGCTTC	3540
AAGGTTGCCA TCAAGAGACT CAAAGCCTGT CTTTGTCAC CTGATAAGAA CTCAATCGGT	3600
GTATTCAAGT GTTCTCAAG ACCATTTCCT ACTTTTCAA TGGTTGCCTG AAATTCATCC	3660
TTATAGCTAG TCAAGCGTCG TGGTAACAAT CCACGCTTTT CACCACGAAA CTTGGCGATT	3720
AAAAGATTTT CAGCGACCGT CATACTGGGA GCTGTCCCA TCTTTGGATC TTGGAAGACA	3780
CGAGACAGGT ACTTGGCAGC CTTCTCGGGT GAAAACCTTAG TGAGATCTTC ACCTAAAATA	3840
CGGATAGTTC CACTAGTTAG TGATAAGGTC CCTGCTATAG TGTAAAGAG AGTTGATTTT	3900
CCAGCACCAT TTCCGCCAA AATCGTGATA AAGTCCCGTT CAAAAATTC TAAGGAAACA	3960
TCATTTAAAA TAATCTTTT TCTATCAAAG CCATTTTAA CGATTTTGGT TGCATTTT	4020
AATTCTACAA TTGCTGTCAT TTGCTTAACT TGGCTCCTTT CAAGATTGTT TGCTTAAATG	4080
TTGGAATCAT GAGGCAGACT GCTAAATCA AGGCACTGTA TAAACGAAG TAACTTGTAT	4140
TAAAGCCAAG TGCATAACT GCCCACACTA AAAATTGATA AGCGATAGAA CCTACAACGA	4200
TAGTAACCAA ACGCTCTGCC AAGCTCAAAC TCTTGAAAT AACTTCTCCA ATAATCAAAC	4260
TTGCAAGCCC CACAACGATA ACCCCGATCC CTCGAGACAC ATCGGCATAA CCTTCTTGCT	4320
GAGCAATGAG GGCACCTGCA AGGGCAATCA CACCATTGTA TAAGACCAAG CCCATGAGCT	4380
CCATGCGTCC AGTATGAATC CCGAACTTC TAGCCATATC AGGATTATCC CCTGTAGCAA	4440
TATAGGCTTG TCCGAGTTTA GTGTCCAAGA AAAAGAGCAT GAGAGCAATA ACAATACTCA	4500
CAAAGATGAG ACCTGTCAAG AGTTGATTCA AATCCGAATC AAAAGGCAAA ACATCCTGAA	4560
TTTGCTTGGT TCCAAGCAGG CCTAAATTCG CACGTCCCAT AATCAAGAGC ATGATTGAGT	4620
GACAAGAAGT CATCACAAA ATCCCTGAGA GCAAGGTTGG GATCTTCCCT TTTGTATAAA	4680
GAAGGCCTGC TGCCATTCCA GCCAAACAAC CTGCTCCTAC AGCAACAAGT GTCGCTAAAA	4740
ATGGGTTTAC GCCTTTGGTT ATCAAAGTGA CAGCAACAGC TCCCCAAGA GGAAGGAAC	4800
CTTCTGTCGT CATATCTGGA AAGTTTAAAA TCCTAAATGT CATAAAGATT CCCAGACCTA	4860
GAATAGCCCA GACAAATCCT TGAGAAATAA TGGAAACAAT CATATTTTAT TTAATCCTTT	4920

1070

CTATATTCAT CTTTTTAAAA AATGGGAAGA GTCTCCTCCT CCCTACCTTA TTTATTCGAT	4980
GACTTGTCCT GCTTCTTTGA GAACAGACTC AGGAATAGTA ATACCTAGTT CTTGTGCTAT	5040
TTTTTTATTG ATGACTGACT TACCAGTTGA AAAGACATTG ACTGGGGTAT CGGCTGGTTT	5100
TGCACCTTTC AAGACTTGCA CAATCATTTT ACCTGTTGCC ACACCAAGGT CATGTTGGTC	5160
AATTACAACAT GATGCCAAAC CACCTACTTC TACCATAGCT GTCGCACTGG GATAAATTGG	5220
TTTCTTAGAA CTTTGATTGC TAGAGACAAC CGTTGGAAAT CCTGATGCAA TGGTGTATC	5280
AATTGGAACC CAAATAGCAT CTACCTTGCT AGTCATAACA GTGACAGTTG AGGCAATTTT	5340
ATTGTTGAA GGAAGTCAA ATGTTCCAC TGTCAGACCT GCCTTTTCAG CATAAGCCTT	5400
AAATTC	5406

(2) INFORMATION FOR SEQ ID NO: 167:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 9711 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

CAGCTTGCTC TTAATAATAT AGCAGATGTT ATAGCTGGAA TTATCTTGTA TTTCTGCTGC	60
AAATGGCTAG ATGGTAAGAA GTAGACCGAA TGACTAGCCT ATAAACACCC GTTAAATCGC	120
TAAGATACGT CAAAAAAGCC CTTAACTATG GCACTAGTTA GGGGCTTTGG TGTCTAATG	180
AACCTTATAC ACTAACTACA TTCTAGCATA TAAGCCCAGA TATTCAAGA GTTTTATTTA	240
TTGTTTAAAG TTCTGAAAGG TCTATAATGA AGTTAGCCAT CTAGTATCAA AAAACCGACT	300
AGCTCTTATG AACTAGTCGA TTTCTCATCA ATGCGCCAAC ATTTCTTGGG CGATTCTCTG	360
GCCAGATAGG TTATCTGGGT AGTAGGTTGG CCAGTTGTCC ATTTCTTCAA AGAGGGCTTC	420
TTGGCTTG TG CCTCAAAGA AGATATGGAA ATGTTCTGCC TTAAGTGGG CAACATTGCG	480
GTCCTAAAC TGAACATACT TGAATTGTCC AGCGTCAGCA TCTGTGGCTT CAAAGAGGAA	540
ACGCACGCCA CGATTGCCTT TCTTGTAAGT CAAAATTTTC TTACCGACAT ACTTGTAAGT	600
GTATTTCTTG CTTTGTCCAC CTTGAACAAA TTCCATAGTA TTATCAGTAA TGTTAATCTT	660
AGTCACATCT GTATGATAGC CTTTGTGATA GTAAGCCTTG TACTCAGCCT GGGTCATCTT	720
ACCAGTCAAC TTAGCCTTGT AGTCAAAGAC TTGGTCAAAC GTGCCGTCTT CAAGGAAAGG	780
ATAAAGTATG TGCCAGTTAC CTGCATAGTC ACTCAAGGTG CGGTCTTGA CAGCTGCATC	840
CTCGAAGTAA CCATTTTGA CTTGCTTGGT ATCCTCTGCC TTTTCAGGTT CAATTGCTGG	900

1071

GCCTTCTTGG TCTGTTGTTT GTTTCAAAGC CTTGAGGTTT TTCTCCATCA CGGAAATGTA	960
GTTTTCTCCA GCCTTGGTGT CCTCTTCTGT CAGACTTTCT AAAGGATTGA GGACATCAGT	1020
TTTGACACCT GCTTCTTTTG AAAGTGTTT AGCAAGGGCT TGTGAGGCAT TTCTTCAAAA	1080
TAGATATAGG CGATTTTATT TTTCTTGACA TACTCTGTCA ATTCTGCCAA GCGAGCAGCT	1140
GATGGCTCTG CATCTGGAGA AAGTCCTGAG ATTGCGACTT GTTTGAGTCC ATAGTCCAAG	1200
GCAAGATAGT TAAAGGCTGC GTGTTGAGTC ACAAAGCTCT TTTGTTTTGC TTGAGACAAA	1260
CCTTCTGCGT AAGCCTTATC CAAGGCTTGC AATTTTTTCG TATAGGCAGC TGCATTCTTC	1320
TCAAAGGTCT CTTTTTATC AGGATAATCT GCTGACAAGC TGTGCGGAT GTGCTCTACT	1380
AGTTTAATGG CACGAACTGG TGATAACCA ACATGGGGT CAACTCATG GTGATGACCT	1440
TCTTCTCCAT GGTCAATGGT TCCCTCTTCT TCCTCGCCAC CTGGCAAGAG CAACATATCG	1500
CCTGTCGCCT TGATGGTTTT CACTTTTTTC TTATCCAAG TATCTAGCAA TTTAGGTACC	1560
CATGTTTCCA TGTTTTCATT TTCATAAAG AAGGTATCTG CATCTTGGAT TTTGGCAACT	1620
GCCTTGGCAG ATGGTTCGTA TTCATGAGGT TCTGTCCCAG CACCGATTAG GAGTTCTACA	1680
TTAGCCGTAT CTCCTGCGAC TTGCTTGGTA AATTCATAGA CAGGGTAAAA GGTGTGCACG	1740
ATATTGAGTT TACCATCTGC CTGTTTTTGA TTGGAACAAG CCACTAAAA CAAGGCACAT	1800
AGACTGGCTA GTAATAAGCT AATTTTTTTC ACGTTCGTCT CCTATTTGAT AAAACGTCTT	1860
ACTAACTGA TTAGTATAAA GACAGTTACA AAAATAATGG TAATACTTGC ACTTGCAGGT	1920
GTTTTCTGCAT AGTAGGAAAT GTAAAGTCCT GCTACCATTC CAAAAAGCC AATCGCACTG	1980
GCAAGCAGCA TAACCGATT AAAGTTTTTC CCCAGACGCA GGGCAATACT AGCTGGCAAG	2040
ACCATAATGG TCGATACCAG AAGAGCTCCT GCTGCAGGAA TCATAAGGGC AATAGCCACC	2100
CCTGTCACCA TGTTAAAAAG AATGGACATG GTACGAACTG GCAAGCCATC CACAAAGGCC	2160
GTATCTTCGT CAAAAGTTAA GATATACATA GGACGAAGAA AGAGAAAGGT CAAAATCAAA	2220
ACAACCGCCG CAATGACAAA GAGGGAAATG ACCTGTTCTT CACTGATAGT CACGATCGAA	2280
CCAAAGAGAT ATTGGTCCAA ACTCATTGAA CTCGAGCTTT TACCCTTGCT CATGACAATC	2340
AGAGAAACAG CCAGACCTGT TGACATGAGG ATAGCTGTCC CGATTTCCAT AAAGCTCTTG	2400
TAAACCGTAC GGAGATACTC CAGAAAGACC GCCGCAATCA AGACAATGGC AATAGTAGAA	2460
ACAGTTGGAG AAATCCCCAA AACCAGACCA AAGGCTACAC CTGAAAGTGA GACGTGGCTA	2520
AGGTATCAC TCATCAAAT CTGACGACGC AAGATGAGGA AGGTTCCCAA TACCGGTGAG	2580
AAAAGACTCA TAGCAATAAC GCCTAAAAAG GCGCGTTGTA TAAAGTCGTA AGATAATAAA	2640

1072

CTAAGCATGG CCCACCTCCT GGCCATTCTC ATGAACATTG AAACAACGCC ATGGCGAGTC	2700
TTGGTTACGG ACTAGATGAA TATTGCGATC CGCATAATCC TTAACCTCTT CAGGGTCATG	2760
GGTAATCATC AAAACAGCCT TGCCATGATG ATGGGCGCTG TGGTGATGA GTTCGTAAAA	2820
TTCATTTTTA CTTCTGCGAT CCATCCCCGT TGTCGGCTCG TCTAGGATAA ACACATCAGG	2880
GTCAGAAGCA AACATACGCG CAATTACCGC TCGCTGCTTT TGTCCCCCAG ATAGAGACCC	2940
CAAGCGTTTG TCTCGATGTT CCCACATGCC AACTGAGTCC AGACTAGCCT TGATATGCTC	3000
CTCATCATGA GCATTCAAAC GACGGAACCA GCCTTTTCTC GGATAGCGAC CCGACTTGAC	3060
AAATTCATAG ACCGTAATTG GAAAACCAGC ATTAATACTG GCAATTTGTT GAGGAAGATA	3120
GGCTATTCTC AATTCTTAC CTTGCGTATT TGTCTTTGAA ATAGCCACCT TTCCAATGCG	3180
TGGTTGCAGA ATTCCAAGAC TAGCCTTGAT GAGCGTCGTC TTAGCCGCTC CATTTTCCCC	3240
AGTCAAGGTA ACAAATTCCC CACTATCAAC ACAATAATTG ATATGTTCAA GAACAGGCTC	3300
CTTATCATAA TAGAAGGACA AATCCTCTAC CGTAATATAT CTCATTATTT GATTTCCTCT	3360
ACTAAAGCAG TCAAAAACCG CTGAATCACT TTTTGTTTCAT TTGGAGTAAA CTGAGTCGCC	3420
ACTTGTTTCAT AGGTTAAAAG TGTATGCTCA TGGTGATGGT GGTGCTCCTC AGCGATTGGA	3480
CGAGCCAAGT CAGTCAACTG ATAAAAATC ACACGCGCAT CTTTAGAATC TTTAGATGTT	3540
TCCAACATCC CTTCTTGAC CAAAGACTTA ATGGCCTTGG TAACTGCCGC CTGACTGACA	3600
TTGAGACGAC GGGCCAATTC TGAATTTGTT AAAGATTCCCT CTGACAAGAG CATAAGGATA	3660
TGCTCCTGAG TATTGGTCAG GGCCACCTCG CTAGTGCAAT GACCTATTAG GATTTCATGC	3720
TGATTTTCCG CCGCAAAAT CACCTCATTC AAAAAAGCAT TGATATCCTT TGCTAGCTGT	3780
CTCATATCTG ACTCCTTTCC TTTTAGACTT CTCTTTTTTA AGAGAAAAAT ACTATTCTTT	3840
GACATTTTGT TTACCAGTTA ATTATATCAC AAGCAAAAAA AGAGTCAAGA AAAACGTGA	3900
AAACTAGTTT CATTCTTGAA CTCTTCTATA TTATATTATC TATTGAAATT CTTTGACATC	3960
TCCATCATAA GTCGCCAAT CTTTGCTGAA AAAGCGCTCA TTCAGATGGT AAGTCGGAGC	4020
TGGTGTGGGA TTGGATAGGA AAGGATCAAC TGCCTTGTC AAGGCCAACC AACCCAACCA	4080
ACCAAGGTGA ATGGTGTCTT TCATAAAGAA AGGCTCCCCG CCGTCCTTAG AAAAATCTGC	4140
TATATTGGTA AAACCTTGAC TTTCTAACTG GTAGCGAATC TTCTGCACCG TTTGTTGGTA	4200
CATATCCTCT CGTAGACCAG CATAGTTCAT CCATTTTTTA TTAACAGGTG GAATGATAAA	4260
AATCGGGTTT ACCTTAGATT TAGAAAACCTG TGTTAAAACC AACTGCAAGT CATTATACTC	4320
TGGCGACTTG AGATAGGTAA AGCTTTTCTG AGAATCCTTT AATTTCTTCA AATCCTTCTT	4380
GATCTGCTCA TTATAGAAAT AATTTTCCAT TCCCATCTCA TTATTGGAAG TATTTTTTTC	4440



1073

AGCATCTGCT TTGACAACAT CTTCTATTGC CTGATAAGAA AACTGGTCTG GCAAGATTTT	4500
TAAATACTTA GCTACATGCT TATCGTAGTT AACATAGCCT CTAACCGAAA ACTGACCAAA	4560
AAAGGAAGCT TGGCGTTCAT TAAAACGAGC CAATAATTCA ATCATTTCAT TGTCTGCTGT	4620
CGACAATTCT TCTTTACTTG CCAACTTCTG AACCAGGTCC TTCATAGCTA CGTTTGGGAA	4680
CTGTTGCAGT AAGCGAGTCG CTGCATATTG ACTAGCCTGA TCCCCAGATT GATGTTTCAG	4740
AAAAC TAGTC AACTGGTCTC CATTAAAATA CTGCTGGAAG GCTGCTGGAT CATAGCCATT	4800
TTTACTGAAC CACTGAGGTG AGATAACATA CACAATTGT TTATTCTCCA GCTGTGGTAA	4860
CATCTGTTGC ATTCCAAAAT ATTGGTTAAG CGATGCAGCT CCCCCCTGTC CTAAAAGATA	4920
AGGACGGTAG GAACGATTGT ATTTCTCAGC TAATACCGCA GGATGAGCAC CGTCAAAACG	4980
AAGCCATTCA CTAGAGCCAA AGAAGGGAAC AAAACGCACA TTTGGATCAG ATAGTGCTCT	5040
GACTTTTGA CTTGCTCCT TAAAACATC GATAGTAGTA GCCACTGCTG AACGCTTTTC	5100
AGCTCCTAGA TTATGATGCA TCTCAGTAGG ATAAAAGAAA ATGAGCAGAA AAACCAACAA	5160
ACCAGCGATC AAGACCGGTC CGAAGATCAT CCATAAGCGT TTAAGCATTT TGTAGCTCCA	5220
CAATACCAGC TATGATTTTA TTAGCTGTAT TCCAGTCGTC ACGACCAAAC TCTGTTACAG	5280
GGACACGAAT GTCAAAACGG TTCTCAATCT CCACAATCAA CTCAACCGTT CCCATACTAT	5340
CCAAGACACC TGCATCAAAA AGATCTTCAT CCATCATGTC AGAAACATCT TCCATAAACA	5400
ACTCATCAAT AATTTCAATA ACTTCTGATT TGATATCCAT ATTTTATTTT CTTTATTTTT	5460
TTAAACCATA GATTATTCAA GAATCCAGAA AAGATTAAGA ATGACAACAT GACAACATGG	5520
AAAGTGACAA CCATGCCAAG CAACTGAATC CAGCGATTCT CAGGTAGGGC AGCCTTCCCT	5580
GCTTTTTTCC GTTCCTTATT GAGCGTTTTT TTCTTGCGAA CCCAGGCATC ATTGATGACC	5640
AAGCCTAGTC CATGAAAGAG TCCATAGGCG ATATAGTACC AGGTCACACC ATGCCAAAAT	5700
CCCATAATCA GCATATTTAC AATGTAGGCC ATGCTTGAGG TTACATTACG ATTTTTAAAG	5760
ACTTCTTTT TGGTTAACAC CATCACCATT CGCATAAAGA CAAAGTCACG GAACCAGAAG	5820
GACAGACTCA TATGCCAGCG ATTCCAAAAC TCCTTTAAAT CCCTTGATAA AAAGGGCTTG	5880
TTAAAGTTGA TAGGGCTACG GATTCCCATC AAGTTTGAGA TGGCCAAAGC AAACATAGAA	5940
TAACCTGCAA AGTCAAAGAA GAGTTCCAGA CCAAAAGTAT ACATAACTGC CAAGGCATAG	6000
AGATTAAAGA AGCCACCTGA CTGCAAGGCT AAATTCTTCA GAGGAGGTAG TAAGGTCTCT	6060
CCTAAAACAT GAGCTAGGAT AAATTATAC AAAAAGCCCC ACATGATATA GCGGACAGAT	6120
TCATCCAGCA TATCCATCAA CTCATCTCGC TCAGGAATAG CCTGATAATT TTCATTAAAT	6180

1074

CGCTTAAAGC	GATCGATTGG	ACCACTCGAG	AAAGTTGGCA	TGAAGAGAAG	GAAACGGAGG	6240
AATTCACAGA	GGGTAAAATC	CTTAATCACT	CCATCTCTCA	GCTCGATGAC	AATTCACACC	6300
GAACGAAAGG	TCAGGTAAGA	AATTCACCAAG	AACCCAAGCA	AAGACTGCGT	TCCATTGATA	6360
GCTGGTTGCA	CCTTGACAAA	GATAATCGGA	AGTAGGGACA	GAAAACTAAC	TAAGTAGAAG	6420
ACCCACTTGC	CATCCTTGCT	TTTTTCGATAA	TGCTTGTAGA	AAAGCAGGAG	CAATATTTCC	6480
CAGCAAAGGT	AAATACCCAA	GGCAGCTAGT	TGATTGGTCT	TTCCACCCAC	CAACATGGTG	6540
ACAATAAAGA	AGAGACTTAC	CAACACTTCA	TACCAGGCAA	AGCGTTTCTT	GAAAAAGAGA	6600
CCTATAAAGA	TGGGCAAGGT	TGCAGCAATC	ACATAAACAA	AATACTGAGG	ATTGCCATAT	6660
GGCTCTAAAT	GAGGAAGCTG	TTGAAAAAAC	TCCATCATCT	CTTATTCACC	TCGTTAATCA	6720
ATCCTTTGAT	GTCAATCTTT	CCATTTGGAG	TTAGTGGCAA	ACTGTCTCGG	TAAAGGAATT	6780
TAGATGGCAT	CATATAGGAC	ATCATGATGT	CTGTCAGGTC	TTCTTTGATG	GCCTTGGTAA	6840
TATCGATATC	TCGCTCAAAC	TGCTCACGAA	CACCGTCTTT	TAAGATGACA	TAAGCCAATA	6900
GATTTTGTAC	CTTGTGGTCC	TTGTTATAGC	GCGGTACTGC	GACAGCAGAT	TCGATAAAGC	6960
GAGACTTGTT	GAGGTTTTGA	GAGACATCTT	CTAACTCAAT	GCGGTAACCG	TTAAACTTAA	7020
TCTGGAAGTC	CATCGCTCCG	CCGTAGAGAA	GCAAGCCCTC	ATCTGTCATG	GTTCCACAT	7080
CGCCTGTGTG	ATAGGCTGGC	AGATCTTCAA	ACTCAAAGAA	GGCTTCTGCT	GTTTTTTCAG	7140
GATTGTTTAT	ATAACCTTTT	GAAACAGCTG	GCCCAGAAAC	AATGATTTCT	CCCTGCTCAC	7200
CATTTGGCAG	TTTATTTCTT	TCCTCGTCAA	TGATAAAGGT	TGGAGAATCA	GCCTTGGTAT	7260
AGCCGATTGG	TAGGCGTTTG	AGAGTCGCTA	ACATCTCGTC	TGTCACGGCA	ACTGCTGACA	7320
GAGCTACTGT	CGCTTCTGTT	GGGCCGTAAG	CATTGATGAT	ACGGGCATTT	GGGAAACGCT	7380
CGCGCAGTTT	TTGAGCTGTT	TTGACCGTCA	ATTCTTCACC	ATCAAAGTAG	AAATGCGTGA	7440
TTCCAGGCAT	TTTCTCACTG	TTGAAGTATT	CAGACAACAT	GGCCATATCT	GCAAAGGATG	7500
GTGTTGATGT	CCAGATAGCG	ATTGGCAATG	AAAAGATAGC	CGCAAAGAGT	TGCTTAAAT	7560
CCTGAGTGAT	GACTGAAGGA	AGAGTGAAAA	GCGTACCACC	AAGTGCCAAG	GTCGGTGCCC	7620
AATACATGAC	AGACAAGTCA	AAAGAATAAG	GTGGCTGTGC	CAGCATTTGC	GGACGACTCG	7680
GTGTCGCAAA	TTCTTTATCC	GTAATCATCC	AGTTTGTAAA	GCTGAGGAGA	TTATCATGTG	7740
AAATCTGCAC	TCCCTTAGGC	TTACCAGTCG	TACCAGAAGT	AAAGATAATG	TAGTAATTAT	7800
CATCTCCCTT	GACTGGATGC	GTGATTTTAT	AGTTATTTCC	TTGGGCAAAG	GCTTCTTGAA	7860
CCTGAGCTAG	ATTTATCATT	GGTGTAGAAA	CCTGCTCCAA	GGGAAAGGCT	GAAATGGCAA	7920
TAATCAAGCT	TGGCTCTGCT	ACTTCTAAAA	TAGCTGAAAC	TCGCTCCAAG	GCCGAATGGC	7980

1075

TATCAATTGG AATGTAGGCA TGACCTGACT TAGTCAGCGC TACAAAGGTT GCCAACATTT	8040
CATATTCTTG GCCACCAAAA ACAACCACAG GAGACTTCTC AGGCAAGCCT AGTTGGTCAA	8100
TGACTGCAGC CAAACTATCC GAATCAGCCT TTAAATCGCC ATAAGTGTGT TCCTGCCCCA	8160
AAACATTATA GACAGGATAG CTAGGCTGTG TCTGAGCAAA ATGCTCAATG GTTCAATCA	8220
TATCTGCTAT TGGTTTATTT GACACAATAG GGATTCTCCT TCAAGTTAAA ATTCATTATA	8280
GATAAAGCTT CCTTGACCCT GACCAAGATA GCTAAAGAAG TAAAGCAGCC CTAGAAAGAT	8340
AAGAAAATAC AAGGCTGTCC GACCAAGAAA GAGGTACAAT TCTTTTCTCT GTTTCATCAA	8400
GAAAAACCAT TCATTTCTGT AATTTTTCGC TAAATAAGA GTGATTCTTA CTAGCTTATT	8460
TTTCTACCAT TGTACCACTT TATATAGTAT CTTTCAATT GTTACCGTA TGTTCCAAT	8520
AGATTTGAGC TTATTTTAAG GATTATACAG TTTTCTATG TATATTTTCA AATAGAGTGA	8580
TCCTGCTTCA AAACCTCCATT TCAGGAGACA ATGAAGTAAA TCTTCCCATA ATAAACACA	8640
CAATATCAAG TTTTTTCAAC ACCTGATACT ATGCGCTTTT CTGATTTTTA AAGACTTTTT	8700
AACCACTCTC TCATTTAAAA TAATCTCGTC TGATATAAAT TAAATAGCT TCTATCATCA	8760
GACAAATGGC TGATAGCCAA AAACCTGATGC TAATACCAAA ACTCTCAGTA ATATAGCTCA	8820
TTAGCAAAAC AAATACTGAA AATGCTAATG TAGAAATCAC TTCAAGAACG GAATAGACAT	8880
TAACTAAATG ATTTTCTCT ACTGTTTCCT GAAGAAATAC ACTTTCAGGA ACTTCTTTTA	8940
GTTGCGATAA CATACCAACT AAAGCTGAAA ATAATAAAAA CATCTGTGCG TTTGGAAAAAT	9000
ATAGAATAGT CAGTGTCACT ATTTCCATAG CTACAAGAGG AAAAGAATA CTTTCCCCC	9060
AAATCATPCA TACCTCTCTC AACTAGATGT AACTTACAAA ACCCCTGACC TCATGAGCCA	9120
CTTCTTCCT CCTCATGAGG TCAGTTTAC TTTCTGCTGT TCCAGTATGG TTTTCTCTCG	9180
CTAGATTTCC TCAAAAGGGC AGACTCCTCC CTTGGTGCGT CACACGATTT TTTTCTCTCG	9240
ACTGTTCTTT AATGCATCAT TAACGACGCT TTTCTTCTAG GTGGTTCATA AGGAACAGGA	9300
AGATTCAGGT TGACTTTTCT AATCCTAGAA TAAAGTGCTG AAAACAATTC GGAATAGGCA	9360
TAGAGACTAG ACAATTTGAG GAGCTGCTTG CGTCTGTTC GAACACATTT TCCCACCACG	9420
TGAAGAAAAA GATGGCGGAA GCGTTTGATT GTTAAAGTTT GGAAGTCACC TCCAGCTAGA	9480
TGTTTGAGAA AAAGATAGAG ATTGTAGGCG ATACAGCTCA TCATCATACG AACTTCGTTT	9540
TTGATTAAGG TTGAACTATC CGTTTTATCG CCAAAAAATC CCTCCTTCAT CTCCTTGATG	9600
AAATTCTCGG CTTGACCACG TCCACGATAA AGCTGAAACT GGTCTTGGCT GTTCCACTCG	9660
TCATATTTGT AACGAGAGAA ATAACATCGT AGAACAAGTA TCCTTCTTTT C	9711

1076

(2) INFORMATION FOR SEQ ID NO: 168:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3025 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

CCCCTTTGTC AAAACTGTAA AATTAACGAC TCAACAATTC ATCTTTACAC CAATCTCAAT	60
GGAAAACAAA AACAAATTGA CCTCTGTCAA AACTGCTATA AGATTATCAA AACAGATCCT	120
AACAATAGCC TCTTCAAAGG TATGACGGAT CTGAACAATC GTGACTTCGA TCCCTTTGGT	180
GATTTCTTCA ATGATCTAAA CAATTTCAGA CCTTCTAGCA ATACTCCTCC TATTCCTCCA	240
ACCCAATCAG GTGGAGGTTA CGGTGGAAAC GCGGTTATG GTTCCCAAAA TCGTGGATCT	300
GCTCAAATC CGCCACCTAG CCAAGAAAAA GGCCTGCTGG AAGAATTTGG TATTAATGTA	360
ACTGAAATTG CCCGTCGTGG AGACATTGAC CCCGTTATTG GCGCGACGA TGAGATTATC	420
CGTGTCATCG AGATTCTCAA TCGTAGAACC AAGAATAATC CTGTCCTTAT CGGTGAACCT	480
GGTGTCGGAA AAACGGCCGT TGTCGAAGGT CTAGCTCAGA AAATTGTCGA TGGCGATGTG	540
CCACATAAAC TCCAAGGTAA ACAAGTCATC CGTCTGGATG TGGTTAGCTT AGTTCAAGGA	600
ACGGGGATTC GAGGACAATT TGAAGAACGC ATGCAAAAAC TCATGGAAGA AATTGCAAAA	660
CGTGAAGACA TCATCCTCTT TATCGATGAA ATCCATGAAA TTGTTGGTGC TGGTTCTGCG	720
AGTGATGGTA ATATGGACGC AGGAAATATC CTCAAGCCAG CCCTTGCTCG TGGAGAACTG	780
CAACTAGTCG GTGCTACTAC CCTCAATGAA TACCGTATCA TTGAAAAGGA TGCTGCCCTC	840
GAGCGTCGTA TGCAGCCTGT TAAAGTCGAT GAACCAACGG TGGACGAAAC AATCACTATT	900
CTCAAAGGGA TTCAAAAGAA ATACGAAGAT TACCACCACG TTCAATATAC AGATGCTGCG	960
ATTGAAGCAG CTGCAACTCT TTCCAATCGC TACATCCAAG ATCGCTTCTT GCCTGACAAG	1020
GCCATTGACC TCCTAGATGA AGCTGGTTCT AAGATGAACT TGACCTTGAA TTTTGTGGAT	1080
CCTAAAGTAA TTGATCAGCG CTTGATTGAG GCTGAAAATC TCAAGTCTCA AGCTACACGA	1140
GAAGAAGATT TTGAGAAGGC GGCCTACTTC CGCGACCAGA TTGCCAAGTA TAAGGAAATG	1200
CAAAAGAAAA AGATCACAGA CCAGGATACT CCTAGCATCA GCGAGAAAAC TATTGAGCAC	1260
ATTATCGAGC AGAAAACCAA TATCCCTGTT GGTGATTGTA AAGAGAAAGA ACAATCTCAA	1320
CTCATCCATC TAGCCGAAGA TCTCAAGTCT CATGTTATTG GTCAAGATGA TGCAGTCGAT	1380
AAGATTGCCA AGGCTATTCC CCGTAATCGT GTCGGACTTG GTACCCCTAA CCGCCCAATC	1440

1077

GGAAGCTTCC TCTTCGTTGG GCCAACTGGT GTCGGTAAGA CAGAACTTTC CAAACAACTG	1500
GCTATCGAAC TTTTGGTTC TGCTGATAGT ATGATTGCGT TTGATATGAG TGAATACATG	1560
GA AAAACATA GTGTGGCTAA GTTGGTCGGC GTCCTCCAG GTTATGTTGG CTATGATGAG	1620
GCTGGTCAAT TAACTGAAAA AGTTCGCCAC AATCCATATT CTCTCATCCT TCTCGATGAA	1680
GTGGAAAAAG CTCACCCAGA TGTATGCAC ATGTTTCTTC AAGTCTTGA CGATGGTCGT	1740
TTGACAGACG GGCAAGGACG CACCGTTAGC TTCAAGGATG CCATCATTAT CATGACCTCA	1800
AATGCAGGTA CAGGAAAGAC CGAAGCTAGC GTTGGATTTC GTGCTGCTAG AGAAGGACGT	1860
ACCAATTCTG TCCTCGGTGA ACTCGGTAAC TTCTTTAGCC CAGAGTTTAT GAACCGTTTT	1920
GATGGCATTG TCGAATTTAA GGCTCTCAGC AAGGATAACC TCCTTCAGAT TGTCGAGCTC	1980
ATGCTAGCAG ATGTTAACAA GCGCCTCTCT AGCAACAACA TTCGTTTGA TGTAAGTAT	2040
AAGGTCAAGG AAAAGTTGGT TGACCTAGGT TATGATCCAA AAATGGGAGC ACGCCCACTT	2100
CGTCGGACTA TTCAAGACTA TATTGAGGAC ACAATCACTG ACTACTACCT TGAAAATCCA	2160
AGCGAAAAAG ATCTCAAAGC AGTTATGACT AGCAAGGGAA ACATTCAGAT TAAATCTGCC	2220
AAAAAGCTG AAGTTAAAG TTCTGAAAA GAAAAATAA TCCTATAAAA AAGGAGTAGA	2280
AAATGAAAT TTTCTGCTTC TTTTTTACT AAAATAACTG TAATTTCTTG ACAGCTTGCC	2340
CTTTGTCCAT TATGATATAT AGTAGACTGA ATCTGAAATA GTACGAAACA ATTGCTAAAA	2400
CATTTATAGA AATTAATTTT ACTTTCCCAA TCGATTTGTT CTCATCTTAT TTCAATCTGC	2460
TATAGTCAAT TGAAACAAGA ACAAGACAAA AGAGCCTCAT AAAAGGTATT GCAACTTGGT	2520
AATACCTTTT TGAGGTGCTT TTTGATATGA GCCCATGTTT TCTCAATAGG ATTGTA CTCA	2580
GGTGAGTAGG GAGGAAGAGG TAAAAGTTTA TACCCTAACT CTTACACAAA GAGTTCTAAC	2640
TTACCCATTC TATGGAATCT TGCATTATCC ATAATAATAA CCGATGGTGT GGTAAATGTT	2700
GGTAAGAGAA ACTTCTGAAA CCAAGCTTCA AAAAAGTCGC TCGTCATCGT CTCTTCGTAA	2760
GTCATTGGAG CGATTAACTC ACCATTCAAT TGTTAGACCT GCAACCAAAG AAATTCTCTG	2820
ATATCTTCTT CCAGATACTT TGCCTCTTCT TAACTGACCT TTTAATGAGC GACCATATTC	2880
TCGATAAAAA TAAGTATCGA ATCCTGTTTC GTCAATCTAA ACAGGTGCTA GGTGCTTTAA	2940
ACTATTAAAA TTCTTAAGAA ATAAGGCTAC TTTTCTGGG TCTTGTTTAT AGTAGGTGTA	3000
GTTCTTTTTT TTTTCGAGTG TAGCC	3025

(2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4104 base pairs

1078

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

TTTAAGGTTT TAAAAAAGT TTTGAAAGG TTTCTTCTTT ATTTTMTAAG GGAGAGATAA	60
CGTTGATATC TAAATCGTGG TCAAAGCCGG CAATTTTTCC TTTAGATGTG TATTGGTGAA	120
TATCATAATC TAAATCAGTT TTAGGACTGC TCTCCAAAAA TCCTGAGTCT GAGCCGTAGA	180
CGGAATCCAA ACAGAGGTAA ACTTGCCTGT ATCAATACTG TGTTCCTCCA TGAAGTAGAC	240
ACCAACGTAG ATGCCGATGT TTTTAGCACC CAGTGATGCT AGTTTGTCTC GAAAGTTTTC	300
GACACCTTCG TTCATATTAG ACATGGTTTT GTCTTCCACG TCAAGCCAAT AGTAACTAGG	360
GCTGTAAGGA GAGGCAGCAT TGTAGAAAAC TTCGGCAGCC TTTCCATTT CTTGGACACT	420
TTTTCCAGCT ACATAAGCGT AGACAGCAAC TGGGACATTC CGCTTTTGAA GTTCAGTGAT	480
ATGACTCTTA TAGGCCTTGT CTATTCCATT GATAAATGAA GCATCATTTT CTTTGTGCT	540
TTGAGCACCA CTGTGAACAC GAACAATAGC ACCTGAAATA TTTGTGAGA GGGCATCGTA	600
GTTGATTTCC TCAGGACGCT GCCAGCCAGA GAGCTCAATA ATCGGTTTGT CTAAGTGT	660
CAAAGCCTGT GCTTCAATCT GTGCTATATT GGATTTTGTT TTAAACGATT GGCTGTCATT	720
AAGTGGGCGA TTGATGATTA AAATGAACAT CATAATCCCA AAAAACTAA ATAAAAAAG	780
TGGATGAATT TGTTTTCTCA TATCTTATAA TTCTACCCTA AAAATCAAAA AAAATCAAAA	840
AAATGGGTTA AGGAAGAGAC TTTAGAGCAT TTTTTCATTC AAGAGTGCGG AATGATTTGA	900
AATATGGTAT AATAAAAGGG AATTTCTACA GAAAAGAGAA GATTATGTCA AATTTTGCCA	960
TTATTTTAGC AGCGGGTAAA GGGACTCGCA TGAAATCTGA TTTGCCAAAA GTTTTGCACA	1020
AGGTTGCGGG TATTTCTATG TTGGAACATG TTTCCGCTAG TGTGGGAGCT ATCCAACCTG	1080
AAAAGACAGT AACAGTTGTA GGACACAAGG CAGAATTGGT TGAGGAGGTC TTGGCTGGAC	1140
AGACAGAATT TGTGACTCAA TCTGAACAGT TGGGAACTGG TCATGCAGTT ATGATGACAG	1200
AGCCTATCTT AGAAGGTTTG TCAGGACACA CCTTGGTCAT TGCAGGAGAT ACTCCTTTAA	1260
TCACTGGTGA AAGCTTGAAA AACTTGATTG ATTTCCATAT CAATCATAAA AATGTGGCCA	1320
CTATCTTGAC TGCTGAAACG GATAATCCTT TTGGTTATGG ACGAATTGTT CGTAATGACA	1380
ATGCTGAGGT TCTTCGTATT GTTGAGCAGA AGGATGCTAC AGATTTTGAA AAGCAAATCA	1440
AGGAAATCAA CACTGGAACA TACGTCTTTG ACAACGAGCG TTTGTTTGAG GCTTTGAAAA	1500
ATATCAATAC CAATAACGCT CAAGGCGAAT ACTATATTAC AGACGTCATT GGTATTTTCC	1560

1079

GTGAACTGG	TGAAAAAGTT	GGCGCTTATA	CTTTGAAAGA	TTTTGATGAA	AGTCTTGGGG	1620
TAAATGACCG	TGTGGCGCTT	GCGACAGCTG	AGTCAGTTAT	GCGTCGTCGC	ATCAATCATA	1680
AACACATGGT	CAACGGTGTT	AGCTTTGTCA	ATCCAGAAGC	AACTTATATC	GATATTGATG	1740
TTGAGATTGC	TTCGGAAGTT	CAAATCGAAG	CCAATGTTAC	CTTGAAAGGG	CAAACGAAAA	1800
TTGGTGCTGA	GA CTGTTTTG	ACAAACGGTA	CTTATGTAGT	GGACAGCACT	ATCGGAGCAG	1860
GAGCGGTCAT	TACCAATTCT	ATGATTGAGG	AAAGTAGTGT	TGCAGACGGT	GTGATAGTCG	1920
GTCCTTATGC	TCACATTTCGT	CCAAATTCAA	GTCTGGGTGC	CCAAGTTCAT	ATTGGTAACT	1980
TTGTTGAGGT	GAAAGGATCT	TCAATCGGTG	AGAATACCAA	GGCTGGTCAT	TTGACTTATA	2040
TCGGAAACTG	TGAAGTGGGA	AGCAACGTTA	ATTTCGGTGC	TGGA ACTATT	ACAGTCAACT	2100
ATGACGGCAA	AAACAAATAC	AAGACAGTCA	TTGGAAACAA	TGTCTTTGTT	GGTTCAAATT	2160
CAACCATTAT	TGCACCAGTA	GA ACTTGGTG	ACAATTCCCT	CGTTGGTGCT	GGTTCAACTA	2220
TTACTAAAGA	CGTGCCAGCA	GATGCTATTG	CTATTGGTCG	CGGTCGTCAG	ATCAATAAAG	2280
ACGAATATGC	AACACGTCTT	CCTCATCATC	CTAAGAACCA	G TAGGAGCCT	ATCATGGAGT	2340
TTGAAGAAAA	AACGCTTAGC	CGAAAAGAAA	TCTATCAAGG	ACCAATATTT	AAACTGGTCC	2400
AAGATCAGGT	TGAATTACCA	GAAGGCAAGG	GA ACTGCCCA	ACGGGATTTG	ATTTTCCACA	2460
ATGGGGCTGT	CTGTGTTTTA	GCAGTAACGG	ATGAACAAAA	ACTTATCTTG	GTCAAGCAGT	2520
ACCGCAAAGC	TATCGAGGCT	GTCTCTTACG	AAATTCCAGC	CGGAAAATTG	GAAGTAGGAG	2580
AAAACACAGC	CCCTGTGGCA	GCTGCCCTTC	GTGAATTAGA	GGAAGAAACA	GCCTATACAG	2640
GGAAATTAGA	ACTCTTG TAC	GATTTTATT	CAGCTATTGG	CTTTTGTAAT	GAGAAGTTAA	2700
AACTATATTT	AGCAAGCGAT	TTGACAAAAG	TGGAAAATCC	GCGTCCGCAG	GATGAGGATG	2760
AAACCTTGGA	AGTCCTTGAA	GTGAGCTTAG	AAGAAGCGAA	AGAATTAATC	CAATCAGGTC	2820
ATATCTGTGA	TGCCAAGACA	ATTATGGCTG	TTCAGTATTG	GGAGTTGCAG	AAAAAATAGA	2880
GGAGGTCAGT	ATGGGTAAAT	CTTTATTAAC	GGATGAAATG	ATTGAAAGAG	CTAATAGAGG	2940
CGAAAAAATT	TCAGGTCCTC	CTTTGCTAGA	TGATAATGAG	GAAACTAAGA	TTTTACCAAC	3000
CTCTTCTTCC	CGTTTTGGTT	ATGCCAATCC	TAAGGATCAT	GGTTTTAGCC	AGGAAACCTT	3060
GAAGATTCAG	GTCGAACCAT	CTATTCATAA	AAGCCGTCGT	ATTGAAAATA	CCAAGAGAAA	3120
TGTCTTCAAT	TCTAAGTTGA	ATAAAATCTT	ATTTGCGGTC	ATCTTTCTCT	TGATTTTGCT	3180
TGTTTTAGCA	ATGAACTTTT	TGTAATAGAA	AAGGAATTGA	AATGAAAATA	GGAATTATTG	3240
CTGCTATGCC	AGAAGAACTG	GCTTATCTGG	TCCAGCATT	AGATAATGCC	CAGGAGCAAG	3300

1080

TTGTTTTTGG GAATACCTAT CATAACAGGAA CCATTGCTTC TCATGAAGTC GTTCTGTAG	3360
AAAGTGGAAT TGGTAAGGTC ATGTCTGCTA TGAGTGTGGC GATTTTGGCT GATCATTTCC	3420
AGGTGGATGC CCTTATTAAT ACGGGTTCAG CTGGGGCAGT AGCAGAAGGT ATCGCTGTTG	3480
GGGATGTCGT GATTGCTGAC AAATTAGCCT ATCATGACGT GGATGTCACA GCTTTTGGCT	3540
ATGCTTATGG ACAAATGGCG CAACAACCGC TTTATTTGCA ATCAGACAAA ACCTTTGTTG	3600
CTCAAATCCA AAAGAGTTTA TCTCAATTGG ACCAAAACTG GCATCTTGGT TTGATTGCTA	3660
CAGGAGATAG TTTTGTGCA GGAAATGACA AGATAGAAGC GATTAAGTCC CATTTCCCAG	3720
AAGTTTTAGC CGTGGAGATG GAGGGGGCAG CTATTGCTCA AGCAGCGCAT GCCCTCAATC	3780
TCCCAGTCTT AGTCATCCGA GCTATGAGTG ACAATGCCAA CCATGAAGCA AACATCTTTT	3840
TTGATGAGTT TATTATCGAA GCTGGACGTC GCTCTGCCA AGTCTTGTG ACCTTTTGA	3900
AGGCTTTAGA TTAAGCGGAA ATTTGACAGT TTTTCTAGCT TATGATAAGA TTTAAGTAAA	3960
GAAAAGCTAG AAAACGTTTC AGAGGATATT ATGAGTATTG AAATGACCGT CAGTGAGATT	4020
GCAGAGGTCT TAGGATTATC TCGCCAAGCA ATCAATAACC GTGTCAAAGA ATTACCAGAA	4080
GAAGACACAG ATAAAAATGA CAAG	4104

## (2) INFORMATION FOR SEQ ID NO: 170:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8876 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

CACGGATAGG CTCGGCTTTC ATCAGTCCTC AGGCTGATTT ACTAATAGCA ACTTTCCTCG	60
ACAAAGTCCA CAGCGATACG TnTGGGTATC AATCCTACGC TTACGCTGAT ACCTTTGCTG	120
GCAGGATTGG CAACGATAGA GCTTGATTGG CTTGGAGTTA CTATTGGGCA AGGATGGTAC	180
AAACCGTAAT CCATCCACTG CTTTCAACAG TTCCTTAAAA TCCCGATCCT TGTGTTGATA	240
GCCTTTCCCT TGAAAATAGA GGTGATAATG ACAGAGTTCA TGTCGGACAA TTTTCCTAAA	300
AACGTCCAAC CCCAGTTCCT GATAAACCTT GGGATTAAAA TCCAAATGCC CATCTTTGGG	360
GAAAAATCGC CCACCTGTCG AACGTAGACG CCTATTCCAC TGGACATGAT GGATAAAAGG	420
TCTGCCGAAG TCTTCTAGTG AAACCTGCTT GACGTAATCA GTCAGTTTCA TTTGGAGCTA	480
GGAGAGACAG ATTAACCTTT TCACGTTTCTG TATCAATTTT CTTAACCCAA ACGCTCACCA	540
AATCTCCAAC TGCCACCACT TGACTAGGGT GTTTGATAAA CTTGCGACTC ATATGGGAAA	600



1081

TATGGATGAG ACCGTCCTCA TGAATCCGA TATCAACAAA AGCACCGAAA TCAACAACGT	660
TACGCACCAC TCCTTCTAGC TTTTGTCCAA CCACTAAGTC CTTGATATCT AGGACATCTT	720
GGCGAaCACA GGTGCGTCAA AGGAATCAGC GAAATCTCGA CCTGGTTTGA GAAGATCTGC	780
AATGATATCT TTAAGAGTTT CTGGACCAAG GTCTAACTCT TGCGCCATTT CCTTGACTGA	840
AAGCGACTTG AGTTTGCTTT GGGCTTCTTC GTTTAGGTCT TTAATATCTA AACGTTTGAA	900
GAGTTCCTTA ACTGCAGTGT AATTCTCTGG GTGAACTCCT GTATTATCAA GGATATTGCT	960
ACTTTCAGGG ATACGAAGGA AACCAGCAGC CTGCTCAAAG GCCTTGGCTC CCAGACGAGG	1020
AACTTTCTTG ATTTGGGCGC GTGAAGTGAT TTTTCCTTCT TCCTCGCGGT ATTTGACAAT	1080
ATTTTCAGAG ATAGTTTGT TGAGTCCAGC TACGTGTGAA AGAAGAGCTG GGCTAGCTGT	1140
ATTGACATTG ACACCAACTT GGTTAACCAC TGTATCGACA ACAAAGTCCA GACTCTCAGA	1200
TAGTTTCTTC TGACTGACAT CGTGTGGTA TTGACCGACA CCAATTGACT TAGGATCGAT	1260
TTTGACCAAT TCCGCAAGAG GATCTTGCAA ACGACGGGCG ATAGAAATGG CAGAGCGTTT	1320
TTCAACGGTC AAGTCTGGAA ACTCCTGACG AGCAAGTTCG CTGGCAGAAT AGACAGAAGC	1380
ACCACTTTCA TTAACGATAA CATAGCTGAC TTCAGGGAAA TCTTTCAGAA CTTCGGCTAC	1440
AAAAGCTTCA CTTTCACGAC TGGCCGTTCC ATTTCCAATG GCAATAATCT CTACACCGTA	1500
TTGACCAATT AAATCTGCTA AATCTTCTT GGGCTTCTCG ATTTGACGAG CTGATGCTGG	1560
TTTAACAGGA TAAATAACCT GAGTTGTCAG CATTTTCTCT GTTGCATCCA CGACAGCTAG	1620
CTTGGCACCT GTACGAAAGG CTGGGTCAA TCCAAGAACC ACGCGCCCTT TCAGTGGAGC	1680
AACCAAGAGG AGATTGCGCA GATTGTCAGA AAAAAGTTGG ATAGCTCCTT CTTCAGCTTT	1740
CTCAGTTAAT TCTGTCCGAA TACGACGCTC GATAGCAGGC AAGACCTTTT TCTTAACGGA	1800
TTGCTGAACA ACTTCATCAA TATAAGCATT TTTACCTTG AAACGAGTAG CAAAGAAGGC	1860
AAGAATACGG TCCGTGCGAT GTTCAAAACC GATCTTCAAG ACACCAAGTT TCTCCCCACG	1920
ATTGAGAGCC AAGGTACGAT AGCCTTGCAAT AGTTCCAAT GTCTCTGAAA AATCATAATA	1980
AATCTGAAAA ACCTGCTTTT CATCAAGACT TTCATCCTTG GCTTGAGAAG TAAGTTTAGA	2040
GTGTCTCAGC ACTTCCTGAT AAGTCATAGA ACGCAAGGTC ACATCTTCCG ATAAGGCTTC	2100
GACCAAAATA TCAACTGCAC CGGTCAAGGC TTCCTTGCCA GTCGCAAATC CTTACAGAC	2160
GAACTTTTC GCTTCTTCT CTAAGTCAAC TATATTCTGC AAAATCAAGC GAGCAAGAGG	2220
AAAGAGTCCA GCTTCACGGG CAATGGTTGC CTTGGTACGA CGCTTTTCTT TATAAGGAAG	2280
ATAGAGTTCT TCAACGTCTG CTAATTTTTC GGCAACTAAG ATAGCTTCTT CCAATTCCTT	2340

1082

GGTCAACTTA CCTTGTCTT GAATCTTAGC TAAGACAGCT TCCTTACGGT CATTGAGATT	2400
TGTCAGACTT TTATCCAAAT CAATAATAGC CTTAATCGCC ACCTCATCCA GACTACCACT	2460
CATGTCCTTG CGATAACGCG CGATAAAGGG AATAGTCGCC CCTTCAGCTG TCAAACCTTAG	2520
AACGGTATCA ATTTGCTTTA ACGTCACTCC CAAATCCTGA GAGATTTTTT CATATTTTTT	2580
ATCCATAAAT CTATTATACC ACAAGCTAAA CGTTTCAAAT TAACTCGTAG AACATTTAAA	2640
AAATATGTAG GAAATAGATT TATATGCTAC AGCGCAATAA CTGCACTTA AAGAGCATTG	2700
CCACCTTTTT TTAACCAAGC CATGATATCA AAAGTATTTA ATGGATCAGA CATAATAGCC	2760
AGTTCTGGAA GATGTTCTCG ACCTGGAATA ACACATTGAC TTTTCAAATT TTTATATGGA	2820
CGATTGACTA AAATTAATTT ATTAGAATAA GGAAGATTAT CCATCTTATT TAAAATTTCT	2880
TCACTAGCTG AATCTTTATT ATCAAATTTA AAATAAAGAT TATTCCAATT TATGCGTTTT	2940
TTTCTTTTTT CCCACTTAGT TCGTGCTTCT TCAATACTAG AATAATGTAG AAAATGAATA	3000
TCTATATCTC CTAAGTGCCC CAAAGGATAA ACTTCATGAG TCCAGCTCGG TGAAATAAGT	3060
TCCTCTTCGA AAACAAGTTC TTGTTCCATA TAATAACGAA AATGCTTTGT AAGTTTATAA	3120
TAATCATCAG GAAGAATAAA TAAACCAACA AAAGGTGTTT TATATTGAAA ACCAAGCTGT	3180
TTATAAATTA ATCCTCCAAC ACAATTATTA CTTATAATCG TAAAATCTAA TCTATCAAGC	3240
TCAAGAAAAG GGAAAATTCC TTTCTCTGCA GCTATTAACT TATGATAAAC AATATCAGAA	3300
TCTAAATATT CACCGTCATT TTTTAACCAA GCACTAAAAT TTGCCAATTC TTGAATATAT	3360
TGTTTTTTCG CTCTTTCTAT ATCATAGTTT TCTAAGACGG CGCAATCTTT GATTCTATTT	3420
TCATAATTTT CTAATATGAT TTGTAGGAG TCTTTTAGAG GTTTAGCATC TATAACAGGT	3480
TTATAGATAT ATGTCGGGAA ATTAATATAG GTTGCACTTT TAGAGTGAAT ATAAAGTCTC	3540
CAAATAAGGT TGTTTATATC AAATTGATTT ATTTTTCGTA AAAGCTTACT ATTGAATAAT	3600
TTTCCAAATA ATGAGCGATA TTGTTTCTA ATTCGATGAT CTGTATCATC CATCTTTTGT	3660
AAAACCTGAA CATTCGTTAA ATTTTCTGTC AACCAATTAT CCCCCAAAA AGGATAAAAG	3720
TAAAATACTC CATCAACCAA ATCAGCAAAA TGACCAAGAA CAACATCAGA ATCGGATAAT	3780
TTTATCGCAT GATACATCTT TTCAAATGTC CAATCAAATA ATGAATCATT TGAAGATAGA	3840
AACGTAATAT AATCTCTGT AATCATATCA GACAACTCAG CAAAAGAATT CTCATCTATA	3900
ATCTTAATAT TAAATGATAG ATTCATCTGT TGGCTAATGG AAGCTATCTC CTCTGTAGAT	3960
TGATTTACAA TAATAACTTC TATATCTTTT AATGTTTGTC TCTCCACTAT TGACAAAGAC	4020
TCTAATAAAC TATTTTATC TCCTTGATGT AACAAAACAA CACTAATTGA GTAAGTCAGT	4080
TTGACTACCT CCCATAATTT TCTGATAATG ATTTTCTTTT TATTTAATTA TAGCACAATT	4140

1083

ATGATATATA TCAGGTAATA TCAAGCTATA TTATCTCTTA GCTACTCAAT TTGAAATTTT	4200
AACTTTTCCC TTTTCCGCAA AATAATAGTA TAATAGAGGT AGAATCTAGA ATCGAGGTAC	4260
ACCTATGGCT GTCAAATTTA CAAAACGAGA CGACTTGGAC AAGATGTTTG AAGAGTTTGC	4320
TAAACTCCCT GATTGAAAC AAGTTACTTT CCCTGATGAC AAAGAGAAAA AAGTCAAAGC	4380
AGAAAAGAAA AACTAGATGA CTGCTTTTCA ACAACTCCCA TCTAGTGAC TTCAAACCTGG	4440
AGCCATTTTT CTCTCCATTA TCATTGAAGC CCTTCCCTTC GTTCTGATAG GAAGCATTGT	4500
CTCAGGGCTG ATTGAAGTTT ATATCACACC TGACAAGGTT TATCATTTTC TCCCTCGAAA	4560
TCGTTGGGGG AGAATCTTTT TTGGGACCTT TGTCGGTATA CTTTTCCCTT CTGTGGAATG	4620
TGGAATCGTC CCCATCATCA ATCGTTTTCT GGAAAAAAG GTTCCAAGTT ACACGGCCGT	4680
TCCTTTTCTT GTGACAGCAC CTGTTATCAA TCCCATGTG CTTTTGCGA CCTATTCTGC	4740
CTTTGGCAAC TCCTTCCATG TCGCCCTATT ACGAGCTCTG GGTTCCTTC TTGTGGCTGT	4800
AATACTAGGA ATTTTTCTAG GATTTTTCTG GCAAGAACCG ATTCAGAAAG AAAATCGTCT	4860
GGCTTGTCAT GAGCATGATT TTTCTTACTT GAGTTCTGCA AAAAAAGTTT TTCAAGTCTT	4920
TGTGCAGGCC ATTGATGAAT TTTTGTATC GGGGCGTTAT TTGGTATTTG GCTGCCTCTT	4980
TGCTTCTATA ATACAGGTCT ACGTTCGAC TCGGATTCTG ACCTCTATCA GTGCGACCCC	5040
TCTTTTGGCC ATCCTGCTCT TGATGATTTT AGCCTTTCTT CTTTCGCTCT GTAGTGAGGC	5100
GGATGCCTTT ATAGGTGCTT CTCTTCTCTC GAGTTTCGGT TTGGCACCAG TTCTGGCCTT	5160
TCTCGTCATT GGTCCAATGC TGGATATCAA AAATATTCTC ATGATGAAAA ATTACTTGAA	5220
AGCACGATTT ATCAGTCACT TCATAACAAT TGTAACCTCT GTCGTCTTAG TCTATTCTCT	5280
CTTGATTGGA GTTATCCTAT GATTCGATTT TTAGTTTTAG CTGGCTATTT TGAACGACT	5340
ATTTACCTCC ATCTGTCGGG CAACTAAAC CAGTACATCA ACATGCACTA TTCCTATCTG	5400
GCCTATATCT CCATGGTGCT TTCTTTTATC TTGGCTATCG TTCAATTGTA TATCTGGATG	5460
AAGCAAGTCA AAACCCACAG TCATCTGAAC AGCCGATTAG CCAAGATAAC GAGTATTTCT	5520
CTTCTGGCTA TTCCACTGT CATCGGCTTA ACTTTCCCAA CTGTTAGCTT GGATTCTCAG	5580
ACTGTTTCTG CTAAAGGTTA TCATTTCCCC CTATCGGAAG GAACGGATCT AGCCATTCAG	5640
ACAAGCGAAG GGACGACAAG CCAATATTTG AAACCAGATA CCAGTTCTTA TTTTCAAAA	5700
TCAGCCTATG AAAAGGAAAT GCGAACGGCG GCGGATAAAT ACTTATCCCA AGATAGTATT	5760
CAGATCACTA ATGAAAATA TATGGAAGTC ATGGAGGCTA TCTACGACTA TCCAGATGAG	5820
TTTGAGGGCA AGACAATCCA GTTTACAGGC TTTGTCTATA ACGACCCAG TCATGCCAAT	5880

1084

AGTCAATTTC TGTTCGATT CGGCATTATC CACTGTATCG CAGATTCTGG TGTCTATGGA	5940
TTGCTGACCA AGGGCAATAC CCGGCAGTAT GAAAACAACA CTTGGATAAC AGCCAAAGGA	6000
AAACTGGTCA ATCACTACCA TAAAGAACTC AAACAAAACC TTCCAACCTT GGAAATCGAC	6060
AGCTTTACCA AAGTCGATAA ACCAGAAAAT CCCTATGTAT ATAGAGCTTT TTAAGAAAAT	6120
CAAGATAAAA ACGAACAAGT TCTCTTCTGA ATAACAGAAA AAGAGCCTGT TCGTTTTTTG	6180
TTATATGAAA ATTAGTGACT TGTAGATTTT CATCTTATAC CATTCCCAGC AATACAAGTA	6240
GCTCATAGAA AATAAGCGAG CCACTCATTC ATTAGACTAG CGATTTCCTT AGGTGCTTGA	6300
GTATAAAGCT CATGGCCAAA GTTTTCTAAA AAAATAGTAT CAAAATAGTC TGGCAATTCT	6360
TTTAGGGCTT CCTCTCTCCA TGTAGCTTCA TTAGGATAGC GAGGACTAAT AAACAAGGTA	6420
TCTCCCCTT CTCTCTTAAA AGCTTGTATT TTTCTCCGTA GCGGAGTATC GCTTCTATAT	6480
TTTCATAATT TATAGCCAAC TCATATCTAT TATACTCAAC ATTCCAGTGA TAAGACTGTC	6540
TTACAGCTTT TTCCATATTT TCTGACCAAT GCTTTGCTTC AGATTTTTCT TTAGAAGTAA	6600
GAACATCTAA GTCCGAAACA ATTTGAGATT TGATATAATT TTTAGTTTCC TCTAACTCTG	6660
TATCCAAAGG TAAAATCTTA TCTAAATCTA GATAGCCACC ATCCAAAAGA ATCAGTTTCT	6720
TTACTTCTTC AAATTCCGAT GCGAAATAAC GAGCTAAATC TCCTCCAAGA GAATGGCCTA	6780
TCAGACAGAT AGATTCTTCC TCTACAAATT CATTTTTAAA CCATGATTTT AATTCTGTTT	6840
CATCTCGAAG ATGCTTTTCA TATGGATTTA GAAAATAGAC CTGCGAATCT AGTTCTTGAA	6900
GAAAATCCTT GCTATGATAG GCATTGCTTC CCAAACCGCC AATAAAATAT TTTTTCATTC	6960
TCTACTTAAT ACTATGCTTA TTCATCTTTT GTTCAAAGAT AGTTGTGATA ATCTGACGCA	7020
ATTCTTCGCG TTTTTTTTCT GGAATCTCAC CACTTGTGTTG AGCTACAGCG TAGAGTTCAG	7080
GGTATTCAAT TGAAATGCGT TTAATCGTAC GTGTTGTAGC ATGTTTTCTG ACAAAAAACG	7140
GGATTGCTT AATCAAGTCT TGTGGGACTA GCGCCAGAAT CTTCTCAGTA GTTCTTTTGT	7200
CACTAATATT AGACATTGTA AGCCTTTTCT TAATCATTTT CTGTTCTTTT TCTGTAAAAT	7260
CTTTTAATTC CATTGATTA GTCCTCCTAT TTTCTCTAAG TTAAATTATG TACTAATACA	7320
GATGAAACTA CAAAGAATAA ACTTTAAGAA ATCTTCTCAC TGATAAGATT TTAGCATTAG	7380
ACTTCCTGCG AAACAAAATA TGGTATAGTA GTTCTATGAA TTATGAAGCA AGTAAACAAC	7440
TAACTGATGC ACGATTTAAA CGTCTTGTTG GTGTTGAGCG CACGACTTTT GAAGAGATAT	7500
TAGCTGTATT AAAACAGCT TATCAACTTA AACACGCAA AGGTGGACGA AAACCTAAAT	7560
TAAGCCTAGA AGACCTTCTT ATGGCCACTC TTCAATATGT GCGAGAATAC CGCACTTATG	7620
AAGAAATTGC GGCTGATTTT GGTATTCACG AAAGCAACTT AATCCGTCGG AGCCAATGGG	7680

1085

TTTAAGTAAC TCTTGTTC	AGTGGTGTTA CGATTTC	AAGTCTCTC AGTTCTGAGG	7740
ACACGGTAAT GATTGATAGC	CATTCCCATC AATATCGTAT	CTTTGGACAT AGCCAATAAA	7800
TGTTTCATTT TTGCGTGGTT	TCTGGCTATT AACGATTGAA	ATAACCCACC AACTTATCAA	7860
AAATAGAAAT AAAAATCCTA	AGATTACTGT CATATCATAA	CACTATTAAA GTTTAACCCA	7920
CTTATCATTA TCCATGATAA	AAGGCTTAGC CAGTCCCTCG	CCTGTATAAT CCGCATACTT	7980
<hr/>			
GGTGCCCAAA TACTTGTAGC	AATCTTCCTT ACTAGCAAAT	TTAATCGCTT GGTAGGGCTC	8040
TTCGAAAGTC AATTTCTCTA	CAAATAAGAA ACCGTCATCA	GCAGGTACTA AGACCCCAAC	8100
GTGGCCTACA AACAGATACT	CGCCATCCAA ATTGTCGTGC	AAGACTACAG ACAGCATTCG	8160
AGCTTTTTC	TTGAATTGAA ATTGTGAGAA	GAATGCTTCC ATCTTTTCAG	CGTGAACCTT 8220
GACATCTGTA GTTGACTCAG	TTGGAACTCT CGAAAATAGA	ATATCAAACCT CTTCTTATC	8280
TTGTGAATCA AAGACCTTTC	CTTTATCAAT CGCATCATTA	TCTAGGAAAA GCAACTGGTC	8340
ATTCTTTTCA AGCTTTGGAA	TGGTGACTGA ATTTTTCAAA	AGACAATAAC TATTGATACG	8400
GCAGTTGGTC CCAACAAAAT	CGCCCTTCTT TTGATTCCAG	AGATGACTGA TTTTCTCAAC	8460
ATCGTATTCG GTGTGAGTAA	AGGAAGTGAA ATCTCCTGAT	AAGCCAGTTG AGCCGACAAT	8520
GGTATTATAG TCATTAACGA	GATTAAAAAA TGCATCAACA	CTATTTGGAT CCAAGTGAGC	8580
TGATAAGAGA GATTTGACCT	CTTCTGTACT TACCTGGTTG	TTAGGTTGG TGTATGAAGC	8640
TTCCATGGA ACTTTCGCTG	AACTGCTTTG CCTTTGATTC	GTCCCCCTCAG AAGTAGCATG	8700
TTGTTGTGTA CAAGCAGCCA	AGCCTAAAAA CAAGGCTGAA	CAGATTCTTA ATGTGGCTAA	8760
TTTTCTTGAT TTCTTCATTT	CTTCTCCTA AATGTCTTGG	ATTAAAGTTT CTTTAACTAT	8820
TGCTTTACAG ATATTGATTA	CTTCTCATT TAATGTGTTT	ATCGTCTTTC CTCCGG	8876

(2) INFORMATION FOR SEQ ID NO: 171:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14736 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

CGCAAACCTT CGCGGTCGGA	AGGTAGTTTT ATGACACGAT	TTGAGATACG AGATGATTC	60
TATCTCGATG GAAAATCATT	TAAGATTTTA TCTGGTGCCA	TTCATTATTT TAGGGTTCCT	120
CCAGAGGATT GGTATCATT	GCTCTATAAC TTGAAGGCTC	TTGGTTTTAA TACGGTAGAG	180

1086

ACTTATGTTG CTTGGAATTT ACACGAGCCT TGTGAAGGTG AGTTTCATTT TGAAGGTGAT	240
CTGGATTTAG AGAAATTTCT CCAAATAGCG CAGGATTTGG GTCTCTACGC AATTGTGCGT	300
CCGTCTCCAT TTATCTGTGC GGAATGGGAA TTCGGTGGCT TACCAGCTTG GCTCTTGACC	360
AAGAACATGC GAATTCGCTC ATCCGACCCA GCATATATCG AGGCAGTTGG TCGCTACTAT	420
GATCAGTTAT TGCCAAGACT GGTGCCTCGT TTGTTGGACA ATGGTGGCAA TATTCTCATG	480
ATGCAGGTTG AAAATGAGTA TGGTTCTTAC GGAGAAGATA AGGCTTACCT GAGAGCGATT	540
CGACAGCTAA TGGAAAGAGTG TGGCGTAACC TGTCCCTCTT TTACATCAGA TGGTCCATGG	600
CGAGCTACTC TGAAGAGCTGG AACCTTAATT GAAGAGGACC TCTTTGTAAC AGGAAACTTT	660
GGTTCTAAGG CACCTTACAA CTTTTCGCAG ATGCAGGAAT TCTTTGATGA ACATGGTAAG	720
AAATGGCCAC TCATGTGTAT GGAGTTCTGG GATGGTTGGT TCAATCGCTG GAAAGAACCG	780
ATTATCACAC GGGATCCTAA GGAATTGGCA GATGCAGTTC GAGAGGTTTT GGAACAAGGC	840
TCTATCAATC TTTACATGTT CCACGGTGGT ACAAACCTTG GTTTCATGAA TGGTTGCTCA	900
GCTCGAGGAA CTTTGGACCT GCCACAAGTT ACGTCTTATG ATTACGATGC CTTTCTGGAT	960
GAAGAAGGAA ATCCAACCTGC TAAATATCTT GCAGTCAAGA AGATGATGCC AACACATTTT	1020
TCAGAGTATC CGCAGTTGGA ACCACTCTAC AAAGAGAGTA TGGAGTTGGA TGCTATTCCA	1080
CTAGTTGAAA AAGTTTCTTT GTTTGAAACC TTAGATAGCT TGTCAAGTCC TGTAGAAAGT	1140
CTCTATCCTC AAAAGATGGA GGAGCTGGGA CAAAGTTATG GCTACCTACT TTATCGAACA	1200
GAAACAACT GGGATGCAGA AGAAGAAAGA CTTCTGATCA TTGATGGTCG AGATAGGGCC	1260
CAGCTGTATG TCGATGGTCA GTGGGTTAAA ACTCAATATC AGACAGAGAT TGGGGAAGAT	1320
ATTTTTTATC AAGGTAAGAA GAAAGGGCTA TCTAGGTTAG ATATCTTGAT AGAAAAATATG	1380
GGGCGTGTC AACTATGGGCA TAAGTTCTTA GCGGATACGC AACGTAAGGG AATTCCGACA	1440
GGGGTCTGTA AGGATCTGCA TTTCTTACTA AACTGGAAAC ACTATCCACT CCCACTAGAC	1500
AATCCTGAGA AAATTGATTT TTCAAAAGGA TGGACTCAAG GACAACCAGC CTTTTACGCT	1560
TATGACTTTA CAGTCGAAGA GCCAAAAGAT ACTTACCTAG ACTTGCTCTGA GTTTGGTAAG	1620
GGGGTTGCCT TTGTCAATGG GCAGAATCTA GGACGTTTTT GGAACGTTGG CCCAACTCTC	1680
TCACTTTATA TCCCTCATAG CTATCTCAAG GAAGGTGCCA ACCGCATCAT TATCTTTGAA	1740
ACAGAAGGTC AATATAAAGA AGAGATTCAT TTAACGTA AACCTACACT AAAACATATA	1800
AAGGGGAAAA ACTTATGACA ATGTAGGAT GCCGTATTGA TGGACGTTTG ATCCACGGAC	1860
AAGTAGCCAA TCTTTGGGCT GGAAAACTAA ATGTTTCACG CATTATGGTT GTAGACGACG	1920
AAGTTGTCAA CAACGATATT GAAAAGAGTG GTTTGAAACT TGCACACCA CCAGGTGTGA	1980

1087

AATTGAGTAT TTTGCCAGTT GAGAAAGCTG CAGCCAATAT TCTTGGTGGC AAATACGATA	2040
GCCAACGTCT CTTTATCGTG GCTCGTAAAC CAGACCGCTT CCTTGGTTTG GTAGAAGCAG	2100
GTGTACCACT TGAAACCCTT AATGTTGGGA ATATGTCTCA AACACCAGAA ACTCGTTCTA	2160
TTACACGTTT TATCAACGTA GTAGACAAGG ATGTGGAAGA CTTCACAAA CTGGCAGAAA	2220
AAGGTGTAA ACTTACTGCT CAGATGGTTC CAAATGATCC AATTTAGAC TTTTGAGCT	2280
TATTAATAA GGAAAAAAT TTTTAGGAGG TCATTGTTAT GATACATGG TGGCAAAATTT	2340
TACTTCTCAC TTTGTACTCA GCTTATCAAA TCTGTGATGA GTTGACGATC GTTTCATCTG	2400
CAGGTTCCTT TGTATTTGCT GGTTCATTA CTGGTTTAAT CATGGGAGAT GTGACTACTG	2460
GTTTACTTAT CGGTGGTAAC TTGCAACTGT TCGTTCCTTG GGTGGGTACC TTCGGTGGTG	2520
CTTCTCGTAT CGACGCAACT TCTGGTGGG TTCTTGCGAC ACCTTCTCTG TTTCACAAGG	2580
AATTGATGCA CCGCTTGCCA TTAACAAT CGCTGTACCA GTAGCAGCTC TCTTGACTTA	2640
CTTCGACGTT CTTGGTCGTA TGAATACTAC CTTCTTCGCT CACCGTGTGG ATGCTGCAAT	2700
CGAACGCTTT GACTATAAAG GTATTGAACG CAACTACTTG CTTGGTGCGA TTCCGTGGGC	2760
TCTATCTCGT GCCCTTCCAG TCTTCTTTGC CCTTGCTTTT GGTGGTGCCT TTGTACAATC	2820
AGTAGTAGAC TTCGTTGAAG CCTACAAATG GGTTCAGAT GGCTTGACAC TTGCAGGACG	2880
TATGCTTCCA GGTCTTGAT TTGCAATCTT GCTTCGTTAC CTTCAGTTA AACGTAACCT	2940
TCACTACCTT GCTATGGGAT TTGGTTTGAC AGCTATGTTG ACTGTTCTTT ACTCATATGT	3000
AACAGGTCTT GGTGGCGCTG TTGCTGGTAT CGTAGGTAAT CTTCTGCTG AAGTTGCTGA	3060
AAAAATTGGT TTCGTGAACA ACTTCAAAGG TTTGTCTATG ATTGGTATTT CTATCGTAGG	3120
TATTTTCCTT GCAGTGCTTC ACTTCAAAA TAGCCAAAAA GTAGCTGTAG CAGCACCTTC	3180
TACACCATCA GAAAGTGGG AAATCGAAGA TGACGAATTC TAATTACAAA CTTACAAAAG	3240
AAGATTTTAA TCAATCAAC AAACGTAGCT TGTTTACTTT CCAATTAGGT TGGAACACG	3300
AACGTATGCA AGCTTCTGGT TACCTTTACA TGATCTTGCC TCAGTTGCGT AAAATGTATG	3360
GTGATGGAAC TCCTGAATTG AAAGAAATGA TGAAAGTTCA TACTCAATTC TTCAATACTT	3420
CACCATCTT CCATACCAT ATCGCTGGTT TTGACCTTGC CATGGAAGAA AAAGATGGTG	3480
TAGGTTCAAA AGACGCCGTT AACGGTATCA AGACAGGTTT GATGGGACCA TTCGCTCCTC	3540
TTGGGGATAC AATCTTTGGT TCACTTGTAC CTGCTATCAT GGGGTCAGTC GCAGCAACTA	3600
TGGCTATCGC TGGCCAACCT TGGGGGATCT TCCTTTGGAT TGCAGTTGCA GTAGCGTATG	3660
ACATCTTCCG TTGGAACAG TTGGAATTTG CTTACAAAGA AGGGGTAAAC CTTATCAACA	3720

1088

ACATGCAAAG TACCTTGACA GCTTTGATTG ACGCTGCATC TGTACTTGGT GTCTTCATGA	3780
TGGGTGCTCT TGTAGCAACA GTGATTAAC TTTGAAATTC TTACAAGTTG CCAATCGGTG	3840
AAAAGATGAT TGATTTCCTAA GACATCTTGA ACCAAATCTT CCCACGTTTG CTTCCAGCAA	3900
TCTTTACTGC CTTTATCTTC TGGTTGCTTG GTAAGAAAGG TATGAACTCT ACTAAAGCTA	3960
TCGGTATTAT TATCGTACTT GCTTTGGCTC TTTCTGCCCT TGGTCACTTT GCACTTGGA	4020
TGTAATTCCT TATGACTAAA TCATTAATTT TGGTGAGCCA TGGTCGCTTC TGTGAGGAGC	4080
TTAGAGGTAG CACAGAAATG ATTATGGGCC CACAAGACAA CATTTACACA GTAGCTCTTC	4140
TTCCAGAAGA TGGCCAGAA GAATTTACTG CTAAATTTGA AGCTGTTATT GAAGGATTGG	4200
ATGATTTCTT AGTCTTTGCG GATCTTCTCG GTGGGACACC TTGTAATGTG GTGAGTCGCT	4260
TGATCATGGA AGGTCGTGAT ATTGACCTTT ACGCAGGGAT GAATCTTCCA ATGGTGATTG	4320
AATTTATCAA TGCGAGCCTT ACAGGCGCAG ATGCGGACTA CAAGAGCCGT GCTGCAGAAA	4380
GCATTGTGAA AGTTAATGAC CTGTTAGCGG GCTTCGATGA TGACGAAGAT GAATAATACT	4440
CTTCGAAAAT CTCTTCAAAC TACGTCAACG TCGCCTTGCC GTAGGTATAT GTTACTGACT	4500
TCGTCAGTCT TATCCGGCAA CCTCAAACG GTGTTTGTAG CTGACTTCGT CAGTCTTATC	4560
CGGCAACCTC AAAGCAGTGC TTTGAGCAGC CTGCGGCTAG TTTCTACAG ATTTTAGTTG	4620
GAAGTCGATT CAATTCATGT GACAACGTGA AAATCGTTAG AGCATTTTAT ATAGAATATA	4680
CATGGGAATG TAGCTTACTC CCATTCCCAT ATTTAATAGA AAAAGAGGAA CTCAATGCTA	4740
CATTATACAA AAGAAGACTT GCTCGAATTG GGTGCAGAAA TCACTACGCG TGAAATCTAC	4800
CAACAGCCTG ATGTATGGAG AGAAGCTTTT GAATTTTATC AAGCAAAACG TGAAGAAATT	4860
GCAGCCTTCC TACAAGAAAT CGCTGATAAA CATGACTATA TTAAGGTTAT CTTGACAGGT	4920
GCTGGGACTT CTGCTTATGT GGGAGATACC TTGCTACCTT ATTTTAAGGA AGTCTATGAC	4980
GAACGCAAAT CGAATTTCAA TGCTATTGCG ACAACAGATA TCGTTGCCAA TCCAGCAACC	5040
TATTTGAAAA AAGATGTGGC AACTGTCCTT GTGTCTTTG CTCGTAGTGG GAATTCGCCT	5100
GAAAGTTTGG CGACTGTGTA TTTGGCCAAA TCCTTGGTGG ATGAGCTTTA TCAAGTGACG	5160
ATTACTGTG CAGCAGATGG TAAATTGGCT CTTCAAGCTC ACGGTGATGA TCGTAATCTC	5220
TTGCTCTTGC AACCAGCTGT CTCTAATGAT GCTGGATTG CCATGACTTC TAGCTTTACG	5280
TCTATGATGT TGACAACTCT CTTGGTCTTT GATCCTACAG AATTTGCTGT TAAGTCTGAA	5340
CGTTTTGAAG TTGTATCTAG TCTTGCCCGT AAAGTTTTAG ACAAGGCAGA AGATGTCAAA	5400
GAGCTCGTTG ATTTAGACTT TAACCGTGC ATCTATCTAG GCGCTGGTCC TTTCTTTGGA	5460
CTTGCTCATG AAGCTCAGCT CAAGATTTTG GAATTAAC TG	5520



1089

TATGAAAGCC CAGTTGGCTT CCGTCACGGT CCAAAATCTC TTATCAACGA CAATACAGTT	5580
GTTTTGGTCT TTGGTACAAC GACAGACTAC ACTCGTAAGT ACGACTTGGA CTTGGTTTCGT	5640
GAAGTTGCTG GTGACCAGAT TGCTCGTCGT GTTGTGCTTT TGACTGATCA AGCTTTTGGT	5700
CTTGAAAATG TCAAAGAAGT GGCCCTTGGT TGTGGCGGTG TCTTGAATGA TATTTACCCT	5760
GTCTTCCCTT ACATCGTTTA TGCCCAACTC TTTGCTTTAT TGACTTCACT CAAGGTAGAA	5820
AATAAACCAG ATACACCGTC TCCTACAGGT ACAGTAAACC GTGTAGTACA AGGTGTCATA	5880
ATTCACGAAT ATCAAAAGTA AGACAGTGTT TATGAATTCT TGACAAGAGG ATTTGTAAAT	5940
TATCAGATAA ACCATAGATT GTCAGTACGC TTTCTATGGT TTGTTTGCTT GAGAGAAATA	6000
GTAAAAGGAG AACAGAATGA AAGCATACAC AGAGCGTGTA TTTGGAAAAT TTGAGGGTGA	6060
GGATGTCTTG GCCTATCGAT TTGAGACAGA CGGTGGCTAC CAACTTGAGG TTATGACTTA	6120
TGGTGCCTAT ATCTTGCCT ATGTCCGACC TGACAAGGCT GGAAATTTTG CCAATGTTAT	6180
CTTGGGATTT GATGACTTTG ATAGTTATGT AGGCAATAGT CCCAAGCATG GAGCAAGTGT	6240
AGGTCCTGTA GCGGGTCGTA TTGCAGGTGC GACCTTTGAG CTCAATGGTA AGACCTATGA	6300
CCTTGAGGTT AATAATGCTA GCAACTGTAA TCACAGTGGT TCAACTGGTT GGGATTCCAG	6360
CTTGTTTGAA GTTGAAGAAG TAAGCGATCA TGGCTTGACT CTCTACACAG AGCGTACAGA	6420
TGGGACAGGA GGGTTCCCTG GAAATCTCAA GATTTGGATC AGTTATCACT TGGAAGAAAC	6480
TGGTGCCTAT GAAATCAGCT ACAAGGTAAC GACCGATCAG GATACGCTGG TCAATCCAAC	6540
CAACCACAGC TATTTCAACT TGTCTGGTGA TTTCACGCAG ACGATTGACC GTCATGTCTT	6600
CCAATAAAC ACAGAGGGCA TTTACTCAAT CGCTCCTGAC GGTGTTCTTG CCAAACTCC	6660
AGAAGCCAAC CGTGATGTGG TCAAACACGT CTACAATGGT ACCTTGTTGA AGGATATCTT	6720
TGCAGAAGAA GATGAGCAAA TCCAGCTGGC ATCAGGTTTG GATCATCCAT TTGCCCTTCC	6780
TGCAGGCCAT GACAATGCTG GATTCCTTTA TGACCAAAAT TCAGGTCGCT TCCTGCTTTT	6840
CAAGACAGAA GCTCCTTGCT TTGTGGTCTA CACAGCAAAC TTTGTGGATG AAAGTGTCAT	6900
CATAGGAGGT CAGCCAATGC TACAGCACAA TGGGATTGCT CTTGAAGCGC AAGCTTTACC	6960
AGATGCCATT CACAGTGACC TTAAAGGCCA AGTCATTCTT AAAGCTGGTC AAACCTTAC	7020
CAGTAAGACA CGTTATGAAC TTGTTGTGAA GTAAAAGAGT CATTCGCGCT ACTTTTGGA	7080
GCTAGGAATA GGTACGCAGA GACAAATAGT AGGAAAATAT GATATAACTA AGCGTTGAAA	7140
GCTATCTGTT AATATAATAT TCAAATACA ATAAGGAGTA AGAAAGAAAC GAAGAAAATT	7200
GTATTTGCTA GTGCCTTGGC TTTGACCTTG GCTGGAGCAG TTTTGACAAA TGATGTTTTT	7260

1090

GCGAACGACA GACTTGTGGC AACACAACT ACTGATGGTA AAAATGAAAA TGTATTGACC	7320
TCAGAGGTGC TAAAACCTTC TAGTGGCAAT GTTTTGGTTG GAATCAAAGG AGAATTTGTG	7380
GCTCCTCATC AACAATCTAT TTTGGATGCC ATCAATGCTA TCTGTAAAGA AGCGGCTGAC	7440
GAAGGTTTGG TAGATAAGTA TGTCCCTATC AAATGATCAA CTGACCTAGA AAAGGCAGCT	7500
TTTGCCAGAG CTACAGAAGC ATCTATAACC ATGGATCATA CCCGTCTTTC TAGCAAAGAT	7560
CTTTGGAGTG CCTTTCCAAC TTCTAATAGT ATAATGGGAG AAAATTTGGC ATGGAATCAT	7620
GACGGTTTTC TAAAAGCTAT TGAACAATGG CGTGCTGAAA AAGCAGATTA TGTGGAGAAA	7680
AAAATAGTGG TTCAGACAAC GGGAAATCTG GTCACTATGA GTCGCTAATT AACCTTAAAT	7740
TTACACACAT GGGGATGGCA GCTTTTAAAA ATCCTAACAA TCAATACAAA GCTATTACAA	7800
TTGCTCAAAC TCTAGGTGAT GATGCTTCTT CAGAGGAATT GGCTGGTAGA TATGGTTCTG	7860
CTGTTCACTG TACAGAAGTG ACTGCCTCAA ACCTTTCAAC AGTTAAAACT AAAGCTACGG	7920
TTGTAGAAAA ACCACTGAAA GATTTTAGAG CGTCTACGTC TGATCAGTCT GGTGGGTGG	7980
AATCTAATGG TAAATGGTAT TTCTATGAGT CTGGTGATGT GAAGACAGGT TGGGTGAAAA	8040
CAGATGGTAA ATGGTACTAT TTGAATGACT TAGGTGTCAT GCAGACTGGA TTGTAAAAAT	8100
TTTCTGGTAG CTGGTATTAC TTGAGCAATT CAGGTGCTAT GTTTACAGGC TGGGGAACAG	8160
ATGGTAGCAG ATGGTCTAC TTTGACGGCT CAGGAGCTAT GAAGACAGGC TGGTACAAGG	8220
AAAATGGCAC TTGGTATTAC CTTGACGAAG CAGGTATCAT GAAGACAGGT TGGTTTAAAG	8280
TCCGACCACA CTGGTACTAT GCCTACGGTT CAGGAGCTTT GGCTGTGAGC ACAACAACAC	8340
CAGATGGTTA CCGTGTAAT GGTAAATGGT AATGGGTAAA CTAGGCTCAG GCCATAGGTA	8400
AAGCATTCTAT CTTACTTAGC AAAAAGAATG AACGATAAGA AAGAGTTGA TGGCGAACAT	8460
TGGCCTCTTT TGATTATATA AGATTGGATT CTTGTGCGCT CAATTCAGA CTTTCTATT	8520
GTAAGCTAAT ATTTTATAGC CCATTAAAAG CATAAGCGGT AATCTAATTT AAAAAATGCT	8580
GTAATTAGTC TGAAGTCCAC ACTTACTTGT TGAGATGTTA TCTCTGTTTT TTATCGTTA	8640
AATTTACTGT ATTTTTTATA GTATGCAGAA TATTTTAAAG TATATTTCAA TAGAAATTTT	8700
TATCGATTTA TTGTATAATG ATAAGTAATT GTTGAAAAGT ACTCAGAAAA TTCCATACTA	8760
TATTATTTTT ATGTTTATAC TTTTATGCTA TAAAATATAG ATTGATATAA AGAATATAGA	8820
AAAAGCGAGG TTAATATGAG CCGAAAAAGC ATTGGTGAGA AACGCCATAG TTTCTCGATG	8880
AGAAAGTTGT CAGTGGGATT GGTATCAGTT ACTGTATCTA GTTTCTTTTT GATGAGTCAA	8940
GGGATTCAAT CGGTATCGGC CGATAATATG GAAAGTCCAA TTCATTATAA GTATATGACC	9000
GAGGGTAAAT TGACAGACCA GGAAAAATCC TTGCTGGTAG AGGCCCTTCC ACAACTGGCT	9060

1091

GAAGAATCAG ATGATACTTA TTA	CTTGTT TATAGATCTC AACAGT	TTTTT ACCGAATACA	9120
GGTTTTAACC CAACTGTTGG TACT	TTTCCTT TTTACTGCAG GATT	GAGCTT GTTAGTTTTA	9180
TTGGTTTCTA AAAGGGAAAA TGG	AAAGAAA CGACTTGTTT	ATTTTCTGCT GTTGACTAGC	9240
ATGGGAGTTC AATGTTGCC GGCC	AGTGCT TTTGGGTGA CCAGCC	CAGAT TTTATCTGCC	9300
TATAATAGTC AGCTTCTAT CGG	AGTCCGG GAACATTTAC CAG	AGCCTCT GAAAATCGAA	9360
GGTTATCAAT ATATTGGTTA TAT	CAAACT AAGAAACAGG ATA	ATACAGA GCTTCAAGG	9420
ACAGTTGATG GGAATACTC TGCT	CAAAGA GATAGTCAAC CAA	ACTCTAC AAAACATCA	9480
GATGTAGTTC ATTCAGCTGA TTT	AGAAATG AACCAGGAC AGG	GGAAGGT TAGTTTACAA	9540
GGTGAAGCAT CAGGGGATGA TGG	ACTTTCA GAAAAATCTT CT	ATAGCAGC AGACAATCTA	9600
TCTTCTAATG ATTCATTCCG AAG	TCAAGTT GAGCAGAATC CG	GATCACA AGGAGAATCT	9660
GTAGTTCGAC CAACAGTGCC AGA	ACAAGGA AATCCTGTGT CT	GCTACAAC GGTGCAGAGT	9720
GCGGAAGAGG AAGTATTGGC GAC	GACAAAT GATCGACCAG AG	TATAAACT TCCATTGGAA	9780
ACCAAAGGCA CGCAAGAACC CGG	TATGAG GGTGAAGCCG CAG	TCCGTGA AGACTTACCA	9840
GTCTACACTA AGCCACTAGA AAC	CAAGGT ACACAAGGAC CCG	GACATGA AGGTGAAGCT	9900
GCAGTTCGEG AGGAAGAACC AGC	TACACA GAACCGTTAG CA	ACGAAAGG CACGCAAGAG	9960
CCAGGTCA TG AGGGCAAAGC TAC	AGTCCGC GAAGAGACTC TAG	AGTACAC GGAACCGTA	10020
GCGACAAAAG GCACACAAGA ACC	CGAACAT GAGGGCGAAG CG	GCAGTAGA AGAAGAACTT	10080
CCGGCTTTAG AGGTCACTAC ACG	AAATAGA ACGGAAATCC AG	AATATTCC TTATACAACA	10140
GAAGAAATTC AGGATCCAAC ACT	TCTGAAA AATCGTCGTA AG	ATTGAACG ACAAGGGCAA	10200
GCAGGGACAC GTACAATTCA AT	TGAAGAC TACATCGTAA AT	GGAATGT CGTAGAACT	10260
AAAGAAGTGT CACGAACTGA AG	TAGCTCCG GTCAACGAAG TC	GTTAAAGT AGGAACACTT	10320
GTGAAAGTTA AACCTACAGT AG	AAATTACA AACTTAACAA A	AGTTGAGAA CAAAAATCT	10380
ATAACTGTAA GTTATAACTT AAT	AGACACT ACCTCAGCAT AT	GTTTCTGC AAAACGCAA	10440
GTTTTCCATG GAGACAAGCT AG	TAAAGAG GTGGATATAG AA	ATCCTGC CAAAGAGCAA	10500
GTAATATCAG GTTTAGATTA CT	ACACCCG TATACAGTTA AA	CACACCT AACTTATAAT	10560
TTGGGTGAAA ATAATGAGGA AA	ACTGAA ACATCAACTC AAG	ATTTCCA ATTAGAGTAT	10620
AAGAAAATAG AGATTAAAGA TAT	TGATTCA GTAGAATTAT ACG	GTAAGA AAATGATCGT	10680
TATCGTAGAT ATTTAAGTCT AAG	TGAAGCG CCGACTGATA CG	GCTAAATA CTTTGTAATA	10740
GTGAAATCAG ATCGCTTCAA AG	AATGTAC CTACCTGTAA A	TCTATTAC AGAAAATACG	10800

1092

GATGGAACGT	ATAAAGTGAC	GGTAGCCCTT	GATCAACTTG	TCGAAGAAGG	TACAGACGGT	10860
TACAAAGATG	ATTACACATT	TACTGTAGCT	AAATCTAAAG	CAGAGCAACC	AGGAGTTTAC	10920
ACATCCTTTA	AACAGCTGGT	AACAGCCATG	CAAAGCAATC	TGTCTGGTGT	CTATACATTG	10980
GCTTCAGATA	TGACCGCAGA	TGAGGTGAGC	TTAGGCGATA	AGCAGACAAG	TTATCTCACA	11040
GGTGCAATTA	CAGGGAGCTT	GATCGGTTCT	GATGGAACAA	AATCGTATGC	CATTTATGAT	11100
TTGAAGAAAC	CATTATTTGA	TACATTAAAT	GGTGCTACAG	TTAGAGATTT	GGATATTAAA	11160
ACTGTTTCTG	CTGATAGTAA	AGAAAATGTC	GCAGCGCTGG	CGAAGGCAGC	GAATAGCGCG	11220
AATATTAATA	ATGTTGCAGT	AGAAGGAAAA	ATCTCAGGTG	CGAAATCTGT	TGCGGGATTA	11280
GTAGCGAGCG	CAACAAATAC	AGTGATAGAA	AACAGCTCGT	TTACAGGGAA	ACTTATCGCA	11340
AATCACCAGG	ACAGTAATAA	AAATGATACT	GGAGGAATAG	TAGGTAATAT	AACAGGAAAT	11400
AGTTCGAGAG	TTAATAAAGT	TAGGGTAGAT	GCCTTAATCT	CTACTAATGC	ACGCAATAAT	11460
AACCAAACAG	CTGGAGGGAT	AGTAGGTAGA	TTAGAAAATG	GTGCATTGAT	ATCTAATTCTG	11520
GTTGCTACTG	GAGAAATACG	AAATGGTCAA	GGATATTCTA	GAGTCGGAGG	AATAGTAGGA	11580
TCTACGTGGC	AAAACGGTCG	AGTAAATAAT	GTTGTGAGTA	ACGTAGATGT	TGGAGATGGT	11640
TATGTTATCA	CCGGTGATCA	ATACGCAGCA	GCAGATGTGA	AAAATGCAAG	TACATCACTT	11700
GATAATAGAA	AAGCAGACAG	ATTCGCTACA	AAATTATCAA	AAGACCAAAT	AGACCGGAAA	11760
GTTGCTGATT	ATGGAATCAC	AGTAACTCTT	GATGATACTG	GGCAAGATTT	AAAACGTAAT	11820
CTAAGAGAAG	TTGATTATAC	AAGACTAAAT	AAAGCAGAAG	CTGAAAGAAA	AGTAGCTTAT	11880
AGCAACATAG	AAAAACTGAT	GCCATTCTAC	AATAAGACC	TAGTAGTTCA	CTATGGTAAC	11940
AAAGTAGCGA	CAACAGATAA	ACTTTTACACT	ACAGAATTGT	TAGATGTTGT	GCCGATGAAA	12000
GATGATGAAG	TAGTAACGGA	TATTAATAAT	AAGAAAAATT	CAATAAATAA	AGTTATGTTA	12060
CATTTCAAAG	ATAATACAGT	AGAATACCTA	GATGTAACAT	TCAAAGAAAA	CTTCATAAAC	12120
AGTCAAGTAA	TCGAATACAA	TGTTACAGGA	AAAGAATATA	TATTCACACC	AGAAGCATTT	12180
GTTTCAGACT	ATACAGCGAT	AACGAATAAC	GTAATAAGCG	ACTTGCAAAA	TGTAACACTT	12240
AACTCAGAAG	CTACTAAAAA	AGTACTAGGA	GCAGCGAATG	ATGCAGCCTT	AGATAACCTA	12300
TACTTAGATA	GACAATTTGA	AGAAGTTAAA	GCTAATATAG	CAGAACACCT	AAGAAAAGTA	12360
TTAGCGATGG	ATAAATCAAT	CAATACTACA	GGAGACGGTG	TAGTTGAATA	CGTAAGTGAG	12420
AAAATCAAAA	ATAACAAAGA	AGCATTATATG	CTAGGTCTTA	CTTATATGAA	CCGTTGGTAC	12480
GATATTAAAT	ATGGTAAAAA	GAATACAAAA	GATTTATCTA	CGTACAAGTT	TGACTTTAAC	12540
GGAAATAATG	AGACTTCAAC	GTTGGATACT	ATTGTCGCAT	TAGGAAATAG	TGGACTAGAT	12600

1093

AACCTGAGAG	CTTCAAATAC	TGTAGGTTTA	TATGCGAATA	AACTTGCATC	GGTAAAAGGA	12660
GAAGATTCAG	TCTTTGACTT	CGTAGAAGCG	TATAGAAAAC	TGTTCTTACC	AAACAAAACA	12720
AATAACGAGT	GGTTTAAAGA	AAATACAAAG	GCATATATAG	TCGAAATGAA	GTCTGATATT	12780
GCAGAAGTAC	GAGAAAAACA	AGAATCACCA	ACAGCCGATA	GAAAATATTC	ATTAGGAGTT	12840
TACGATAGAA	TATCAGCACC	AAAGTTGGGG	CATAAGAGTA	TGTTATTACC	ACTACTAACT	12900
TTACCTGAAG	AATCTGTGTA	TATTTTCATCG	AATATGTCTA	CACTTGCATT	CGGTTTCGTAT	12960
GAAAGATATC	GTGATAGTGT	GGATGGAGTT	ATTCTTTTCAG	GAGATGCTTT	ACGAACTTAT	13020
GTAAGAAATA	GAGTTGATAT	AGCAGCGAAA	AGGCATAGAG	ACCATTATGA	TATTTGGTAC	13080
AATCTTCTTG	ACAGTGCTTC	AAAAGAAAAA	CTTTTCCGTT	CTGTGATAGT	TTATGATGGA	13140
TTCAATGTAA	AAGATGAGAC	AGGAAGAACT	TATTGGGCAA	GGTTAACGGA	TAAAAACATC	13200
GGCTCTATTA	AAGAATTCTT	CGGACCTGTT	GGGAAATGGT	ATGAGTATAA	TAGTAGTGCA	13260
GGAGCGTATG	CGYATGGAAG	TTTAACGCAC	TTTGTGTTAG	ATAGATTATT	AGATGCTTAT	13320
GGAACGTCGG	TTTATACTCA	TGAAATGGTT	CATAATTCTG	ATTCTGCAAT	CTACTTTGAA	13380
GGAAATGGTA	GACGTGAAGG	ATTGGGAGCG	GAGTTATACG	CACTTGGTTT	ACTGCAATCT	13440
GTAGATAGTG	TAAATTCTCA	TATTTTAGCT	TTAAATACGT	TATATAAAGC	AGAAAAAGAT	13500
GATTTGAATA	GATTGCATAC	ATATAATCCG	GTGGAACGTT	TCGATTCCGA	TGAGGCGCTT	13560
CAAAGTTATA	TGCATGGATC	ATATGATGTA	ATGTATACAC	TTGATGCGAT	GGAAGCAAAA	13620
GCGATATTAG	CTCAAAATAA	TGATGTTAAG	AAAAAATGGT	TTAGAAAAAT	AGAAAAATTAT	13680
TACGTTCTGT	ATACTAGACA	TAATAAAGAT	ACACATGCAG	GAAATAAAGT	CCGTCCATTA	13740
ACAGATGAAG	AAGTAGCTAA	CTTAACATCG	TTAAACTCAT	TAATCGACAA	CGACATCATA	13800
AATAGACGTA	GCTATGATGA	TAGTAGAGAA	TATAAACGAA	ATGGCTACTA	TACTATAAGT	13860
ATGTTCTCTC	CTGTATACGC	AGCGCTAAGC	AATTCGAAAG	GTGCTCCTGG	AGATATTATG	13920
TTTAGAAAAA	TAGCTTATGA	ATTACTTGCG	GAAAAAGGTT	ATCACAAAGG	ATTCTTACCT	13980
TATGTTTCTA	ATCAGTACGG	AGCAGAAGCA	TTTGCCAGCG	GAAGCAAAAC	ATTCTCATCA	14040
TGGCATGGAA	GAGATGTTGC	TTTAGTGACA	GATGATTTAG	TATTTAAGAA	AGTATTCAAT	14100
GGTGAGTACT	CATCATGGGC	TGATTTCAAA	AAAGCAATGT	TTAAACAACG	TATAGATAAA	14160
CAAGATAATC	TGAAACCAAT	AACAATTCAA	TACGAATTAG	GTAATCCTAA	TAGTACAAAA	14220
GAAGTAACTA	TAACAACGGC	TGCACAAATG	CAACAATTAA	TTAATGAAGC	GGCTGCGAAA	14280
GATATTACTA	ATATAGATCG	TGCAACGAGT	CATACCCAG	CAAGTTGGGT	GCATTTATTA	14340

1094

AAACAAAAA TCTATAATGC ATATCTTCGC ACTACAGATG ACTTTAGAAA TTCTATATAT	14400
AAATAAGATT GTAGAGTTTC ATTGTTGAGT AGTGTTTCTT GTAAGGATGA GGAGTCAGAT	14460
GACAAATCGA CTCCTTTTTC TTATGGATCG ATGTAGAGAT TTGATTGAAT GCAGATTGCA	14520
GGAATCATCT TCAACTCATC AACGACCAAT GGTGACAAGG TGGATTTCAA TCCACAGAA	14580
AATGTTGATT TGAGAAATAA CTTTGCTAGT CTAGTAAAAT AAATACAAAA CAATCCTAGA	14640
AGATTTTTTC TGGGATTGTT TTTTGCTGAG TGGGATGCTT CAAGTGTCTT GGCTTGACTT	14700
TCTTGAGGGA AGTTATATAA TAGTTGTAAT AATTAG	14736

(2) INFORMATION FOR SEQ ID NO: 172:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11770 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

ACAGGAAAGC ACGATAGCAA TCTCTTTGGA AGATTTAAAA AATATTCCTC AAAGTTTCGC	60
TGTTGCTTAC GGTGATACGA AAGTATCTTC GATTCTCTCT GTCTTGCGTG CTAATTTAGT	120
AAATCATTTG ATTACAGACA AAAATACAAT TTTAAAAGTT TTGGAAGAAG ATGGGGATTT	180
GACTTTTAGA GAGATTCTAG GTGAGTGAAA ATGATAGACT GATTCAGTTT ATCGTTTTC	240
TTTTTAGTTG ATTGCACATT TGTGCTTATA TAAACAAAAA TAGTTTATCT GTTGTTTTTG	300
GATTGACAAC TTTATTATGT AGTTGTATTC TATAGTTACA AAAGAAAATT TTAATAATTC	360
AAATGAAAAA AGCTTTTTAC ATAGTGAAAT GAGGAGGAAT TTATGGAAAT GATTGTTCCA	420
GATCAAAATTA TCATGGGTTT AATTTTATAT GCTGGTGATG CGAAACAACA TATTTATAAA	480
GCGTTAGATT ACATAAAAAA TGGTACATGT GAACGGTG TG AAGAAGAAAT ACAGTTAGCT	540
GATGCAGCCT TATTAGAAGC TCATAATCTA CAAACAAAAT TTTTGGCACA GGAAGCGTCT	600
GGTACAAAGA CAGAAATTAC AGCTCTCTTT GTTCATTAC AAGATCATCT CATGACCAGT	660
ATGACGGAGA TTAATTTAAT CAAAGAAATT ATTAGTTTGA GAAAAGAACT TCATAAAAAA	720
TAATACTAGA GTATTATCAT TGTTATTAAC ATAGAGGAGG AAAACATAAT GGTGAAGATT	780
GGTTTGTTTT GTGCAGCAGG TTTTCTACT GGTATGCTTG TAAATAATAT GAAAATTGCA	840
GCGCAATCTA GTGGAGTTGA GGCAGAAATA GAGGCGTTTT CTCAGTCTAA ATTAGCGGAT	900
TATGCGCCAA ATATAGATGT TGCACATTG GGTCCACAAG TTGCTTATAC ATTAGATAAA	960
TCAAAAGAAA TTTGTGATAA GTGTGATGTT CCGATAGCTG TTATTCCGAT GATGGACTAT	1020

1095

GGTATGTTAG ATGGGAAAAA AGTATTAGAT TTGGCCCTAT CTTTGATTAG TGGTAAGAA	1080
AAGGAGATTT ATTATGTCAA AGATGGATGT TCAGAAAATC ATTGCACCGA TGATGAAGTT	1140
TGTGAATATG CGTGGCATTAG TAGCTCTAAA AGATGGGATG TTAGCAATTT TGCCATTGAC	1200
AGTAGTTGGT AGTTTGTCTT TGATTATGGG ACAATTGCCG TTCGAAGGAT TAAATAAGAG	1260
CATTGCTAGT GTTTTTGGAG CTAATTGGAC AGAGCCGTTT ATGCAAGTAT ATTCAGGAAC	1320
<hr/>	
TTTTGCTATT ATGGGTCTAA TTTCTTGTTT TTCAATTGCC TATTCTTATG CTAAGAATAG	1380
CGGAGTAGAG GCTTTACCAG CTGGAGTTCT ATCTGTATCT GCATTCTTTA TTTTGCTAAG	1440
ATCATCTTAT ATCCCTAAAC AAGGTGAGGC GATTGGGGAC GCTATTAGTA AAGTTTGGTT	1500
TGGAGGCCAA GGAATTATCG GTGCTATCAT TATAGTTTGT GTAGTAGGAA GTATTTATAC	1560
CTTCTTTATA AAGAGAAAAA TTGTTATTAA GATGCCAGAA CAAGTTCCAC AAGCTATTGC	1620
CAAACAGTTT GAAGCAATGA TTCCAGCATT TGTAATTTTC TTATCTTCTA TGATTGTATA	1680
TATTTTAGCG AAGTCATTGA CTAATGGCGG AACATTGATA GAAATGATTT ATTCTGCTAT	1740
TCAAGTTCCG TTGCAAGGTT TAACTGGATC TTTGTATGGT GCTATTGGAA TTGCATTCTT	1800
TATATCATTT TTGTGGTGGT TTGGTGTTC TGGGCAATCG GTAGTAAATG GAGTAGTGAC	1860
AGCTCTGCTT TTATCTAATC TTGATGCTAA TAAAGCTATG TTAGCCTCTG CTAATCTATC	1920
ATTAGAAAAT GGTGCACATA TTGTTACTCA ACAATTTTTC GATTCATTTT TAATCTATC	1980
AGGTTTCAGG ATTACGTTTG GTCTTGTAAGT TGCCATGCTT TTTGCAGCAA AATCAAAACA	2040
ATACCAAGCC TTAGGAAAAG TTGCAGCTTT TCCAGCAATA TTTAACGTAA ATGAGCCAGT	2100
TGTATTTGGA TTTCCGATTG TCATGAATCC AGTTATGTTT GTACCTTTCA TTCTTGTTCC	2160
TGTACTTGCA GCTGTGATAG TATATGGAGC TATTGCAACA GGTTCATGC AGCCATTCTC	2220
AGGGGTAACA TTGCCTTGA GTACACCAGC TATTTTATCA GGATTTTGG TGGGTGGATG	2280
GCAAGGAGTT ATTACTCAGC TGGTGATATT AGCGATGTCT ACATGGTTT ATTTTCCATT	2340
CTTTAAAGTA CAGGATCGTT TAGCTTACCA AAATGAAATC AAACAATCTT AGAGGTATTT	2400
GTGTGTTACT GTTAAACTCA CACATTTGTG CTAAAAATTA GAGAGTTAAA ATTTTCTAG	2460
TTAAAGCTT GAAAATTTCT ATAAAAATCG GTATTATATT TTCGAAAGAA ATAAAAATAT	2520
TTTCGAAAGA AAGGTGCTTA CGATGGTAAA TACAGAAGTA GCAAGAACAA CAATCAAGAC	2580
AGAATATTTT GGCAGCCTTA CTGAAAGGAT GAACAAATAT CGAGAAGATG TTTTAAATAA	2640
AAAACCTTAT ATTGATGCTG AGAGAGCAGT TCTAGCAACA CGCGCCTATG AACGATACAA	2700
GGAACAACCT AATGTCCTAA AACGTGCATA TATGCTGAAA GAAATTTTGG AAAATATGAC	2760

1096

TATCTATATT	GAAGAAGAAT	CTATGATTGC	GGGAAATCAA	GCTTCTTCCA	ATAAAGATGC	2820
TCCTATTTTT	CCGGAATATA	CGCTAGAATT	TGTTCTCAAT	GAGTTGGATC	TTTTTGAAAA	2880
GCGTGATGGA	GATGTTTTCT	ATATTACAGA	AGAAACAAAA	GAACAACTTA	GAAGTATTGC	2940
TCCGTTTTTG	GAAAATAATA	ATTTACGTGC	TAGAGCTGGT	GCCTTATTAC	CTGAAGAAGT	3000
GTCTGTTTAT	ATGGAAACAG	GATTCTTCGG	TATGGAAGGT	AAGATGAATT	CTGGAGATGC	3060
TCACCTTAGCA	GTAACTATC	AGAACTTTT	GCAATTGGT	TTAAGAGGTT	TTGAAGAGCG	3120
GGCTCGTAAA	GCAAAAGTAG	CTCTAGATTT	AACAGATCCA	GCAAGTATTG	ATAAATATCA	3180
TTTTTACGAC	TCTATATTTA	TCGTAATCGA	TGCTATTAAA	GTATATGCAA	AGCGCTTTGT	3240
TGCTCTTGCT	AAAAGTTTAG	CCGAAATGC	AAATCCTAAA	CGTAAGAAAG	AATTACTTGA	3300
GATTGCAGAT	ATTTGCTCTA	GAGTCCCAT	TGAACCGCA	ACTACTTTTG	CAGAAGCTAT	3360
TCAATCAGTT	TGGTTATTTC	AATGTATTTT	ACAAATGAA	TCTAATGGCC	ACTCTCTTTC	3420
ATATGGCCCT	TTTGATCAAT	ATATGTATCC	ATATATGAAG	GCTGATTTAG	AAAGTGGTAA	3480
AGAAACAGAA	GATAGCATTG	TTGAACGTCT	GACAAATCTT	TGGATTAAAG	CAATTACAAT	3540
TAATAAGGTT	CGCAGTCAAT	CACATACATT	TTCTTCAGCA	GGAAGTCCTT	TATATCAAAA	3600
TGTTACAATT	GGTGGACAGA	CTCGAGATAA	GAAGGATGCT	GTTAACCCAT	TATCTTATTT	3660
GGTATTAAAA	TCAGTTGCAC	AAACCCATCT	ACCGCAACCT	AATCTAACTG	TACGTTACCA	3720
TGCAGGTTTA	GATGCTCGTT	TCATGAATGA	GTGTATTGAA	GTGATGAAAC	TTGGTTTTTG	3780
TATGCCTGCA	TTTAATAATG	ATGAGATTAT	TATTCCTTCT	TTTATTGCAA	AAGGAGTATT	3840
GGAAGATGAT	GCTTATGATT	ACAGTGCCAT	TGGATGTGTT	GAAACGGCAG	TTCCAGGGAA	3900
ATGGGGCTAT	CGTTGCACAG	GTATGAGTTA	TATGAACTTC	CCTAAGGTTT	TACTTATCAC	3960
GATGAATGAT	GGAATTGATC	CGGCTTCGGG	TAAACGGTTT	GCACCAAGCT	TTGGTCGTTT	4020
TAAGGATATG	AAGAAGTTTT	CTGAATTAGA	AAATGCTTGG	GATAAAACAC	TAAGATATTT	4080
GACACGAATG	AGTGTTATTG	TTGAAAATTC	TATTGATTTA	TCATTGGAAC	GAGAAGTTCC	4140
TGATATTCTA	TGTTCAGCAT	TGACTGATGA	TTGTATTGGT	CGTGGAAGAA	ACCTTAAAGA	4200
AGGTGGAGCA	GTATATGATT	ATATATCAGG	ATTGCAAGTT	GGAATTGCAA	ATTTGTCCGA	4260
TTCATTAGCT	GCAATTAAAA	AATTGGTGTT	TGAGGAAGAA	CGTATAAGCC	CAAGTCAGCT	4320
TTGGCATGCA	CTGGAAACAG	ATTATGCCGG	AGAAGAAGGT	AAGGTCATT	AAGAAATGTT	4380
GATTCATGAT	GCACCTAAGT	ATGGTAATGA	TGATGATTAT	GCTGACAAAT	TGGTTACTGC	4440
TGCTTATGAC	ATTTATGTTG	ATGAAATTGC	TAAATATCCT	AATACACGTT	ATGGAAGAGG	4500
GCCTATTGGA	GGAATTCGTT	ATTCAGGAAC	ATCTTCTATC	TCAGCCAACG	TAGGGCAGGG	4560



1097

ACGTGGAACA	TTAGCAACTC	CAGATGGACG	CAACGCCGGT	ACACCGTTAG	CAGAGGTTG	4620
TTCACCATCA	CATAATATGG	ATCAACACGG	CCCTACATCT	GTTTTAAAAAT	CTGTTTCAAA	4680
ATTACCAACA	GATGAAATCG	TAGGTGGGGT	TCTCTTAAAT	CAGAAAGTAA	ATCCTCAAAC	4740
GTTAGCCAAA	GAAGAAGATA	AATTAAAACT	AATTGCTTTG	TTACGAACAT	TCTTTAATCG	4800
TTTACATGGG	TACCATATTC	AATACAATGT	TGTTTCCAGA	GAGACGCTGA	TTGACGCTCA	4860
<hr/>						
GAAACATCCT	GA AAAACACA	GAGACTTAAT	TGTTCTGTGT	GCAGGATACT	CTGCATTCTT	4920
CAATGTTCTT	TCTAAGGCAA	CCCAAGATGA	CATTATAGGA	CGTACTGAGC	ATACTTTGTA	4980
AAATAAAGAG	GTTCTTTTTA	TGGAATTTAT	GCTTGACACA	TTAAATTTAG	ATGAGATTAA	5040
AAAGTGGTCT	GAAATTTTGC	CGTAGCTGG	GGTAACTTCA	AATCCCACTA	TTGCAAAAAAG	5100
AGAGGGTTCT	ATTAAATTTT	TTGAACGAAT	CAAAGATGTA	AGAGAATTGA	TTGGCTCTAC	5160
ACCCTCTATT	CATGTTTCAGG	TGATTTCTCA	AGATTTTGAA	GGCATCTTAA	AGGATGCTCA	5220
TAAAATTCGA	AGACAAGCAG	GAGATGATAT	ATTTATCAAA	GTACCTGTTA	CTCCAGCTGG	5280
ATTACGTGCA	ATAAAGGCGC	TAAAAAAGA	GGGCTACCAT	ATCACTGCAA	CAGCTATTTA	5340
TACAGTTATT	CAGGGATTAT	TAGCTATCGA	AGCAGGAGCG	GATTACCTAG	CTCCATATTA	5400
TAATAGAATG	GAAAATCTGA	ACATTGATTC	AAATTCTGTC	ATTCTGCAAT	TAGCTCTTGC	5460
TATTGATAGA	CAGAACTCTC	CTAGTAAGAT	TTTAGCTGCA	TCCTTTAAAA	ATGTAGCACA	5520
AGTAAATAAT	GCTTTAGCTG	CAGGTGCGCA	TGCTGTTACA	GCAGGAGCGG	ATGTTTTTGA	5580
ATCAGCTTTC	GCCATGCCAT	CTATCCAAAA	GGCGGTTGAT	GATTTTTCTG	ACGATTGGTT	5640
TGTTATTCAA	AATAGTCGTT	CCATTTAGAT	AGAGAGGAAA	TACATATGAG	AATTTTTGCT	5700
AGTCCTTCTA	GATATATTCA	GGGGGAAAAT	GCCTTGTTTG	AAAATGCCAA	ATCAATTTTG	5760
GATTTGGGAA	ATTGCCCTAT	TCTATTATGC	GATCAGTTGG	TTTATGATAT	TGTTGGAAAA	5820
CGATTTGAAG	ATTACCTACA	TAGGTATGGT	TTCCATATTG	TTCTGGCGCT	ATTTAATGGT	5880
GAAGCTTCTG	ACAATGAAAT	CAATCGAGTT	GTTGCCTTGG	CTGAGAAAGA	AAATTGTGAT	5940
AGTATTATCG	GTCTTGGTGG	GGGAAAGACG	ATTGATAGCG	CAAAAGCTAT	TGCAGATTTG	6000
ATTGAAAAGC	CTGTTATTAT	TGCTCCAACA	ATTGCATCGA	CCGACGCACC	TGTATCTGCT	6060
TTATCTGTTA	TTTATACAGA	TGAAGGTGCA	TTTGATCATT	ATCTATTTTA	TTCTAAAAAT	6120
CCAGATTTAG	TTTTGGTTGA	TACAAAAGTT	ATTTCAACAAG	CCCCTAAGCG	TTTATTAGCG	6180
TCTGGTATTG	CAGATGGTTT	AGCAACTTGG	GTTGAGGCGC	GTGCGGTTAT	GCAGGCAAAAT	6240
GGAAAACTA	TGTTGGGACA	ACAGCAAACA	TTGGCTGGAG	TTGCAATTGC	GAAGAAATGT	6300

1098

GAAGAAACGC TGTTCGAGA TGGTTTACAG GCTATGGCAG CTTGTGAAGC TAAAGTGGTG	6360
ACACCAGCAT TAGAAAATAT TGTGAAGCT AATACTTTAT TGAGTGGTCT AGGTTTTGAA	6420
AGTGGAGGAT TAGCTGCGGC GCATGCAATT CATAATGGTT TTAAGTGCATT GACAGGTGAC	6480
ATTCATCATT TAACACATGG TGAAGAAAGTA GCTTATGGAA CTTTAGTACA ACTATTATG	6540
GAAAATAGAC CTAAAGAAGA ACTTGATAAG TATATTGAGT TTTACAAAAA AATTGGTATG	6600
CCAACAACCTC TAAAGAAAT GCATTTGGAT CAAGTTGGAT ATGATGATTT AATAAAAGTT	6660
GGTAAACAAG CAACTATGGA GGGTGAGACA ATTCATCAGA TGCCGTTTAA GATTTGCGCT	6720
TCAGATGTTG CTCAAGCTAT TATCGCTGTA GATGCCTATG TAAATTCAAA ATAAACAATA	6780
AGGACTACTG TTTTCCAAAT GGTAGTCTTT TATGATCCC TGTATTGAAT TCTATAGAAG	6840
ATTGAAATAG GATGAGAACA AATCGATTGG GAAAGTAAAA TTAATTTCTA TAAATGTTTT	6900
AGCAATTGTT TCGTACTATT TCAGATTCAG TCTACTATAT GTTCTTCATA AATCAAAAAG	6960
CGACATAGGT TGTCGGCTAT TTATTGTGAA TACATTAATT AGCATTCCAG TTTTATCTTC	7020
GGTCTAAAAT AAGTATTTTG TGCTATACGA GATAAGCTTC TTGACTTACT CCTTGATTTA	7080
CTGCATAACA ATGGGATAAA AAGTGGGAGA TAGAGCAATT CATAGTCATC AAAATTAATG	7140
AGATACAGTA TACAGTTTTT CCTTTAAACA CATTTCAAAT TCCCTCAAAA ATGGTATAAT	7200
AGTAACATCA CAAAATTGGA GAGAGACCAT GAGTTTTTAC AATCATAAAG AAATTGAGCC	7260
TAAGTGGCAG GGCTACTGGG CAGAACATCA TACATTTAAG ACAGGAACAG ATACATCAAA	7320
ACCTAAGTTT TATGCGCTTG ATATGTTCCC TTATCCGTCT GGAGCTGGTC TGCACGTAGG	7380
ACACCCAGAA GGTTATACTG CAACCGATAT CCTCAGTCGT TACAAACGTG CGCAAGGCTA	7440
CAATGTCCTT CACCCAATGG GTTGGGATGC TTTTGGTTTG CCTGCAGAGC AATACGCTAT	7500
GGATACTGGT AATGACCCAG CAGAATTTAC AGCGGAAAAC ATTGCCAACT TCAAAACGTCA	7560
AATTAATGCG CTTGGATTTT CTTATGACTG GGATCGTGAA GTCAACACAA CAGATCCAAA	7620
CTACTACAAG TGGACTCAAT GGATTTTCAC CAAGCTTTAC GAAAAAGGCT TGGCCTATCA	7680
AGCTGAAGTG CCAGTAAACT GGGTTGAGGA ATTGGGAACT GCCATTGCCA ATGAAGAAGT	7740
GCTTCCTGAC GGAACCTCTG AGCGTGGAGG CTATCCAGTT GTCCGCAAAAC CAATGCGCCA	7800
ATGGATGCTC AAAATCACGG CTTACGCAGA GCGCTTGCTC AATGACTTAG ATGAACTAGA	7860
TTGGTCAGAG TCTATCAAGG ATATGCAACG CAACTGGATT GGTAAATCAA CTGGTGCCAA	7920
TGTAACCTTC AAAGTAAAG GAACAGACAA GGAATTTACA GTCTTTACTA CTCGTCCGGA	7980
CACACTTTTC GGTGCGACTT TCACTGTCTT GGCTCCTGAA CATGAATTAG TAGACGCTAT	8040
CACAAGTTCA GAGCAAGCAG AAGCTGTAGC AGACTATAAA CACCAAGCCA GCCTTAAGTC	8100

1099

TGACTTGGCT CGTACAGACC TTGCTAAAGA AAAAACAGGG GTTTGGACTG GTGCTTATGC	8160
CATCAACCCT GTCAATGGTA AGGAAATGCC AATCTGGATT GCAGACTATC TCCTTGCTAG	8220
TTATGGAACA GGTGCGGTTA TGGCTGTGCC TGCCACGAC CAACGTGACT GGGAAATTTGC	8280
CAAACAATTT GACCTTCCAA TCGTCGAAGT ACTTGAAGGT GGAAATGTCT AAGAAGCTGC	8340
CTACACAGAG GATGGCCTGC ATGTCAATTC AGACTTCCTA GATGGATTGA ACAAAGAAGA	8400
CGCTATTGCC AAGATTGTGG CTGGTTGGA AGAAAAAGGC TGTGGTCAGG AGAAGGTTAC	8460
CTACCGTCTC CGCGACTGGC TCTTTAGCCG TCAACGTTAC TGGGGTGAGC CAATTCCAAT	8520
CATTCAATGG GAAGATGGAA CTTCAACAGC TGTTCTGAA ACTGAATTGC CGTTGTCTT	8580
GCCTGTAACC AAGGATATCC GTCCTTCAGG TACTGGTGAA AGTCCACTAG CTAACCTGAC	8640
AGATTGGCTT GAAGTGACTC GTGAAGATGG TGTCAAAGGT CGTCGTGAAA CCAACACTAT	8700
GCCACAATGG GCTGGTTCAA GCTGGTACTA CCTCCGCTAT ATTGACCCGC ACAATACTGA	8760
GAAATTGGCT GATGAGGACC TCCTCAAACA ATGGTTGCCA GTAGATATCT ACGTGGGTGG	8820
TGCGGAACAT GCTGTACTTC ACTTGCTTTA TGCTCGTTTC TGGCATAAAT TCCTCTATGA	8880
CCTCGGTGTT GTTCCGACTA AGGAACCAT CCAAAAAC TCCTAACCAAG GGATGATTTT	8940
GGGAACAAGC TACCGTGACC ACCGTGGTGC TCTTGTGGCA ACCGACAAGG TTGAAAAACG	9000
TGATGGTTCC TTCTTCCATG TAGAAACAGG GGAAGAGTTG GAGCAAGCCG CAGCCAAGAT	9060
GTCTAAATCG CTCAAGAAGC TTGTTAACCC AGACGATGTG GTGGAACAAT ACGGTGCCGA	9120
TACCCTTCGT GTTTATGAAA TGTTTATGGG ACCACTCGAT GCTTCGATTG CTGGTTCAGA	9180
AGAAGGTTTG GAAGGAAGCC GTAAGTTCCT TGACCGAGTT TACCGTTTGA TTACAAGTAA	9240
AGAAATCCTT GCGGAAAACA ATGGTGCTCT TGACAAGGTT TACAACGAAA CAGTCAAAGC	9300
TGTTACTGAG CAAATTGAGT CTCTCAAAT CAACACAGCT ATTGCCCAAC TTATGGTCTT	9360
TGTCAATGCT GCTAACAAGG AAGATAAGCT TTATGTTGAC TATGCCAAAG GCTTTATTCA	9420
ATTGATTGCA CCATTTCAC CTCACTGGC AGAAGAACTC TGGCAAACAG TCGCAGAAAC	9480
AGGTGAGTCA ATCTCTTATG TAGCTTGGCC AACTTGGGAC GAAAGCAAAT TGGTTGAAGA	9540
TGAAATTGAA ATTGTCGTCC AAATCAAAGG AAAAGTTCGT GCCAACTCA TGGTTGCTAA	9600
AGATCTATCA CGTGAAGAAT TACAAGAAAT CGCTTTAGCT GATGAAAAAG TCAAAGCAGA	9660
AATTGACGGT AAGGAAATCG TGAAAGTAAT TGCGGTACCG AATAAACTCG TTAATATCGT	9720
CGTTAAATAA CGAGTTTATT AGCTCTATCT GCCACCTTCA ATAGTCCACT GGACTATTGA	9780
AsCCAACATAA ATTAGTTAAC ATTGTTGTGA AATAAGATAG GAGTCCTTCA GAGTAGAATC	9840

1100  
TGGAGGATTT TTTGAATCTT CTTATGAAAG TATGATATAC TATGGGCAAC TATAAAGTTT 9900  
GAAAAGTGAA ATAAGGAGAA TAAGATGCCA GTAAATGAAT ATGGTCAAAT GATTGGGGAG 9960  
TCAATGGAAG CTTATACTCC AGGTGAATTG CCTTCTTTTG ATTTCTTAGA AGGCGTTAT 10020  
GCTAGGATAG AGGCTCTTTC AGTGGAAAAG CATGCGGAGG ATTTATTAGC TGTTTATGGC 10080  
CCTGATACGC CTCGGGAGAT GTGGACCTAC CTCCTTCAGG AGTCAGTAGC AGACATGGAG 10140  
GAACTGGTCA GCCTTTTAAA TCAGATGTTG GCTCGTAAGG ACCGTTTTTA TTATGCAATC 10200  
ATAGACAAGG CAACTGGTAA GGCTTTGGGA ACTTTTTCCC TCATGCGAAT TGATCAGAAT 10260  
AACCGAGTAA TAGAAGTGGG AGCTGTCACT TTTTCTCCAG AGCTCAGGGG GACACGGATA 10320  
GGAACAGAAG CCCAGTATCT CTTGGCTTGC TATGTCTTTG AGGAGCTTAA CTATCGTCGC 10380  
TATGAGTGA AATGCGATGC TCTTAACCTG CCATCCAGAC GAGCAGCGGA ACGTTTGGGA 10440  
TTTATTTATG AAGGAACCTT CCGTCAGGCA GTGGTTTATA AGGGGCGTAC AAGAGATACG 10500  
GATTGGTTGT CTATGATTGA TAAGGACTGG CCTCAAGTCA AAGCTCGATT GGAAATATGG 10560  
TTGCGTCCTG AAAACTTTGA TAAAAATGGA CGACAGCACA AGAGCTTGAG AGAACTTTAA 10620  
GAGGTGTTGA GATGATTACT ATTAAAAAGC AAGAAATTGT CAAGCTAGAG GATGTTTTGC 10680  
ATCTCTATCA GGCTGTCGGT TGGACAACT ATACCCATCA AACAGAGATG CTGGAGCAGG 10740  
CCTTATCTCA TTCATTAGTA ATTTATCTGG CACTTGATGG TGATGCTGTG GTGGGCTTGA 10800  
TTCGTTTGGT TGGAGATGGT TTTTCATCAG TTTTGTACA GGATTTGATT GTTTGCCTA 10860  
GCTATCAGCG TCAAGGGATT GGTAGCTCCT TGATGAAAGA GGCTTTAGGA AATTTTAAAG 10920  
AGGCCTATCA AGTCCAGCTG GCGACAGAAG AGACAGAAAA AAACGTGGGA TTTTATCGTT 10980  
CTATGGGCTT TGAAATCTTA TCCACCTATG ACTGTACAGG AATGATTTGG ATAAACAGAG 11040  
AAAAATAAAA AAACCTGTTT GTTCTTAAGC AAAGTTTAAG GATGGTCTAG TATCATATAG 11100  
TCATTAAATA AAGACCTCCT AACTTTATTT AATAAAATCC TAACTTTTT TCATCACAAT 11160  
CTCCTAATGA AGCCACCCAA TCAGGTGGCT TTTTGCCTG ACGACGGGCA TGTCGTATAT 11220  
CTGAGGTGTA AGTCCTCAGC CTGACTATCG TGAGGTAGCA GGGAGAGGAA GGGATAGCGA 11280  
AATCGTGGCT CTACGAACAG GAACGTGATA GTAAGGCGTA TATAGCGGAT AAGGAGGCTT 11340  
CAAACCTAA AGTCCAAAAA GGTAGTCGTA ACCTATATGT GTAAATCACG AGAGTAATTG 11400  
AATTCGGACT AAGGTTTGTG TGAAAAAGAT AAATCTTTCT AGAGTCTAAA GACTCTGCGT 11460  
CAGATTTCTT ATTTTCACTG TAACCTTTTA ACGTCTCAT ATCTTGATA AACGAGGAAA 11520  
GATGTACGAC TTATCCCGTG AGGTTTCATG AGCGCTGAAA GCGTAGTAAC AACGAATCAT 11580  
GAGAAGTCAG CCGAGCCCAT AGTAGTGAGG AAACCTCCGT AATGGAAGTG GAGCGAAGGG 11640

1101

GTGAATACTC AAACAGTCTG GGGAGAGACT GTTTGAGGTC TGTCGCTAGA AAGAGAAAAC 11700  
 GACAGATCGA AGTAATCCTA CTTCACTTGT GTCTGTAAAA TGAGTGGTCT GATAGAACTG 11760  
 GACTTTGAGG 11770

## (2) INFORMATION FOR SEQ ID NO: 173:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4185 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

CGCGAAACTA CTTTCTTAGT ATAACACTTT CAGAATCATT GTCAATAGAA ATGACTTGAT 60  
 TTTTCAATT TTTCAAGCT ATTCCAAGG GTTGTAATAT CGTCCCTGAT TCTGCAAGAT 120  
 AAGTAGTAAA CTAACACTA AAAACAAGG TGCCAAGAGC AAGGTAATAT AGTCTCCTTT 180  
 TTTCAAGGCC TGATAACTAT ACCATGTGCG TTTTTCTCT TTTCCAAAGC GGCGAACTCC 240  
 ATGGCAGTCG CAATGGTATC AATGCGTTCT AGCGAGCTAA AAATCAAGGG CGTAATAATG 300  
 AGCAGATTGC CTTTGATTGG TTGCATAAGA GAAGCTTTCT TGGATAATTC CATCCACGCG 360  
 GCCTCCTGAG ACATCTTGAT AGTAAAGAAT TCTTCTGCA AATCTGGAAT ATAGCGCAAG 420  
 GTCAGGCTGA CAGAATAAGC AATCTTATAG GGCACACCAA TTTGATTAA ACTGGAAGCA 480  
 AACTGACTAG GATGGGTTGT CATCAAAAAG ATAATAGCCA GAGGAATGGT GCAAAGATAC 540  
 TTAATGGCCA AATTTAGCAG ATAAAAGAGC TCCTGGCTGG TTAGAGTGTA GACACCGATT 600  
 CCCTGCCAAA TCACACTTCT CTCTCCATAA AGTCCAACCC CATACTCGGG AGAAAAGAGA 660  
 TAGACCATCA AAACGTTTAA AACGGCAAAT ATCGTCGCAA AAACGGCTAC AAAGGAAACA 720  
 TCTTTAAAGC GAATTTCTGA TAAATAGAGG AGAAAGACTG AAAAGATGGC AATCAGCAAG 780  
 AGCATTCTGG TATCATAGCT AATCATGGCC GCCAATGATA CCAGAATGAA AAAGAGAAGT 840  
 TTCCCAGCTC CTGACAAGCG ATGAATCACA GTATCTCTAT GCTGGTAACC GATTAATTTA 900  
 GCTTGCATCC CTCTCTCCTT TCTTTGTAAA ATGCCGTTAA ATCCAGTGGA TCCACATCTA 960  
 GTTCTTAGC CAAGTTAAAG ATGGAGGTTT CTTTGTAGATT GGCTTTTACT AACAGCTCAG 1020  
 GATCGCTCAA CAGACTGGCT GGAACAGTAT CGGCAATCAA TTCTCCATCC ACCATGACAA 1080  
 GGACCCGGTC TGAATAATCC AGCATCAATT GCATATCATG GGTAATCATG ACAATGGTAT 1140  
 GCCCTTTTTG ATGTAACCTC TCGAGAAATT CCATAATCTC AGTATAGTTC TTCTGATCTT 1200

1102

GACCTGCAGT CGGTTTCATCT AGGAGAATAA TTTCAGCTCC TAAGACCAAA ATTGAAGCAA	1260
TGGTGACACG TTTTCTCTGA CCAAATGACA GGGCAGAAAT AGGCCAATTA CGGAATTCAT	1320
AAAGTCCACA GATTTTCAAG GTTTCATATA CTCTCGTTTC AATTTCTTTC TCATCCACAC	1380
CTCGCAAACG GAGCCCTAGA GCCACCTCAT CAAAAATCAT ATTGGTTGAA ATCATTTGAT	1440
TAGGATTTTG TAGCACATAT CCTACTCGTT CCGCCCGCTC TGCAACAGAA TCGCCTTTTA	1500
TATCCTGTTT TTCCCAAAGA TAGCGTCCTT CCGTCTGAAT AAAGCTACTT ATAGCCTTGG	1560
CTAGAGTTGA TTTCCCTGCT CCATTTTTTC CGACAATAGC AATCTTTTCA CCCTTTTTAA	1620
TATCTAAATG TAGGGATTTT AAAATCGGTC TATCATCATA AGAAAAAGAT ACTTCCTCTA	1680
GTCTAAAGAG TGA CTGCAAT GCTGGGGTTT CTMTTGCCAG TTCATTCTGC AACTGAACCT	1740
GACCTTTTGA GATAGACAAG TTATCCAGAT TCGCTAATTG TTCTTCCTTG ACTAAGTCCA	1800
CACCTAATTG ACGGAGAGTC GTTAGATAAA GGGGTTCTCG AATTCATTT TGAGTCAATA	1860
AATCAGTCGC AAGCAACTGG TCAGGGCTCC CATTAAAAAG GATACGACCA TCGTTTATCA	1920
AGACAATCCG ATCCACAGGG CGATGCAGAA CGTCTCCAA ACGGTGCTCG ATAATAAGAG	1980
TCGTCGTCCT CTCTTCCTTA TGAATCTGGT CAATCAATTC GATAATATCC TGACCTGACT	2040
TGGGATCTAG ATTGGCGAGT GGCTCATCAA ACAAGAGAAT CGGACTTTCA TCAATCAAGA	2100
CACCAGCCAG ACTGACTCGC TGCTTTTGTC CACCTGACAA ATCCTGAGGA CGCTGATCCA	2160
GTAAAGGAAG AAGGTCCAGC TTTTCAGCCC ATTTATAAAC ACGACCTTTC ATCTCATCTA	2220
GGGCTGTAC ATCATTTTCC AGAGCAAACG CCAAATCTTC TGCCACAGAC AAGCCAATAA	2280
ACTGCCCATC TGTATCCTGC AAAACTGTGC TAACCAGATG AGACTTATCA TAGATGCTCA	2340
TATCAAAGGC TACTTGACCC TTTATCAAAA ATTCTCCATA TGTCTGACCC TTGTAAATAT	2400
TGGGAATAAT CCCATTCAAA CACTGACCCA AGGTAGATTT ACCTGACCCA GATGGTCCAA	2460
CAATTAAGAC TTTCTCTCCC TTGTAAATGG TCAAGTCTAT CCCTTGCAAG GTCGGTTCTT	2520
GTTGTGTTTC ATACCGGAAA GAGAAATCCT TCCACTCAAT TaTAGCTTCT TTCATCTTAC	2580
TCTCTTCATT CGCTTCTTAG ACTTCTATTT TATCATAAAT CAAGCCCTTC TTGCAGTCTC	2640
TCCTCTTAAA ATCTTAGCGC CAAAAAGATT CCTATCCTAG CTTACTTGCC TAACTAATCT	2700
ATAAACATCG AAAAAGACTA GTTGCCAGC CTTCCTCATC ATTTTATACT CTTGAAAAAT	2760
CTCTTCAAAC CACGTCAGCT TCGCCTTGCC GTAGGTATGG TTACTGACTC CGTCAGTTTC	2820
ATCTACAACC TCAAAACCAT GTTTTGAGCC TGCTTCGTCA GTTCTATCCA CAATCTCAAA	2880
ACACTGTTTT GAGCAACCTGC GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAT	2940
TAGTCCTTTT TCAAACTTCC TGCACGAGTT TGGGTTCTCT CATAGGCAAG TAAGAGAAGA	3000

1103

GTTCTGCAA TAGCTACAGA TACACCATTG GCAATTCCCG CAACAATCCC TTGTGCAAAT	3060
ACTTTTTCTG CCGCTTCTTG ATAAATCACA ACATCTCCAA GTGGTGCCAA GACACCCCAA	3120
ACAAGGGCAT TTGCAAGTAG TTGAATGAGA TTAATAATAA GAATATCTTT CCAGTCAAAA	3180
ACACCATTGA TCACGCGAAC GTACTTTCTA AAAAGTCCCA CAACTAAACC AAAGAGTCCG	3240
CTAGCGATAA TCCAAGTCCA CCATAGACCA TAACCAACAA GAGAGTCCTT GATTGCATGA	3300
CCAATCAACC CGACAAGCAA ACCGATAATC GGTCCAAAAA TAATAGAAAG TAGCGCTTGT	3360
ACCGCATACT GAAGCTGGAT GCTTGTATTT GGAACAGGGG TTGGAATGTT GATCATCCCG	3420
ATGACGACAA AGAGGGCAGC GCCAATCCG ACAGCAACAA CTTGTTTAAT TGTAATTTG	3480
ATTTCCATAC TATTCTCCTA TTTTATCCTT CTATTTTCTT TATTTCAATG GTCCAAGATG	3540
AACCGACACC TACATTATAG GCCTTGGCAA AGGAACCTTG GTTGATAGCC AAACCTAAAC	3600
GATAGAGAGA GTTGATGTAA AGGATGGGTT GCCCAATTCT CACATCTGCA AATGATTGTC	3660
CATAGACAAC CTGATTTTGA TAGACCAGCA TATCAGCATG ATAGATGGTC ACTTCAAAAC	3720
GATCACCAAA TTCTGGTTCC AGCTTGTAAG ATTCTTCCCG TGTGATAGAG GTCCAAAGCG	3780
AACCGAAACG CACATCCAGA ATATCAATGG CTCCCTTCAC CAGATGATCT TCTATGATGG	3840
TCGCTACGAC TGGAAGTCT ACAATCTGTT CCACACTGAG CTCTGGCCCT ACTTCCTCAA	3900
AAGTAATGTG ACCACTGGCC AGTTTAGCAC CAGTATAGGC ATAGACATCA CGACCGTGGA	3960
AGGTATAAGA ATGCTCTGTG TTTTGACGCC TATTGGCCAC CTCAGAAATC TCACGAATGG	4020
CTACAATGCC AACGTGTTTC TTGATAAAGG AAAGCGTCCC ATTATCTGGC GTGACAATGT	4080
ATTGATTTTT TGCAGTCTTG GCAACTACAC TCTTACGTTT CGAACCGACA CCTGGATCGA	4140
CAACCGATAC AAACGTCGTT CCCTCAGGCC AGTAATCCAC CGTCT	4185

(2) INFORMATION FOR SEQ ID NO: 174:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2069 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

TGATAGAGTT AAAGCCGCTG AGTCATTCAA TCCATCTCCA ACCATCAAAA TAGTGTGACC	60
TGCTTTCTGC AGTTTCTCTA CTAACCTAAA TTCCCATCA GGTTCAGT CTGTATAGAC	120
CTGATCAAAG GGCAAATCTT TGACTAATTC CTCTGTCCTA ATCAAGGTGT CTCCTGTTGC	180

1104

CAGAATCAAT TTTTCCCCT GTGCCTTAAG TTTATCCAAG GCTGTTTTTG CTTCTTTTCT	240
CAAAGGAGTA TGAATGCAGA ACATTCCAAT CAATTCAATTT TGATAAGCCA AGAATAAGAG	300
ATTGTAGTGA CTCTTGACT CTTCAATTAA AGCATTMTGT TCTGAACTGA TATGAATCTG	360
CTCATCTGC ATCAAGACAT AATTCCTCAAT AAGAACTGGT TGGCCATCTA TATGAGATTT	420
GATCCCCCTG CTGCGATAT ATTGGAGTTT CCCATGCATT TCCTCATGTT CAATTCCCTC	480
TATCTCAGCT TGCTTGACGA TGGCATTAGC AATAGGATGA TAAATGTGTT CCTCAAGACA	540
GGCACTGATT CTGAGAATAT CTTCTCACT ATAGTCTCCA AAAGGTAACA CCTTTTCAAC	600
TATAGGATAA CTACTTGTA TTGTTCTGT CTTATCAAAC AAGAAAGTAT CAACTTCCAG	660
ATATTTCTCC AGAACATCTC CATCCTTAAT CACCATTTC ACGTTCAACC CTTCTTGAT	720
AACTGTCAAA TAAGCTACAG GAGTAGAGAT TTTCAAAGCG CAGGAGAAAT CGACCAATAG	780
GAAAGAAATA GCCTTAGAAA AAGAACCTGT CAATAGGTAA GTCAGCCCAG CCCCCAAGAA	840
ATTATATTTG ACGACTTTAT CCGCCATCTT GATGAAATAG CGTTGTTTCG TTTTCTTGTT	900
TTCTTCAGAT TTCTTCATCA ACTCAATCAG CTGTAAAATA CGGCTGTCA TCTGATTATC	960
TGTTACACGA ATGCGTAACT CTCCAGTTTC TAATACTGTA TTTGCACAAA CCAATCAGA	1020
CTCTCTTTTT TCAACTGGAA AACTCTCTCC TGTCAAGGAA CTTTCGTTGA CCATACCTAA	1080
ACCTGAAACT ACTTGTCAT CAAACAGAAT TTCATTTCTT TGAGATAAGA TCAAGACATC	1140
TCCTATTTGA ACATCGGAAC TCTTGATACT AACAAACGTA TCGCCCTGTA CTAGGAATAC	1200
ATCGCTCTCT TTTGCAAGAA GACTCTGTTT TAAATCTGTT GCAGTTTTTT TCAAGGACCA	1260
CTGATCTAAA TGATTCCCCA AATCAAGCAT AAACATGATA TTGCTAGCTG TCTTGGATTG	1320
GTTCATAAAC AAAGACAATA AAATAGCCGA ACAGTCCAAG ACTTCCATCG TTAGTYCCTT	1380
ACGCGCTAGT GTTTGATAGG CTTCTCTAAT ATAACCCAAA GCCTGATAAC AAGTCCATAT	1440
ATAGCGAATA GGATACGGCA CAAACTACG AAAAAGTACA CGCTTAACCG CTGCACCTGA	1500
AACAATAGAA TAAGCACTCT CTTCTCTACG AATGGGAAGA GTCATCAACT CAGAACTTT	1560
CCCTTTATCA ATTCTTTTTA AAAAGGCTTC TGCATTATCT AATACAGAAA AGCCTTCTTT	1620
TATGCGTAGA GTAAAGTGCT GTTGATCCAT GTAAACTGG ATAGACTCAA TCCCCTTTTC	1680
ATCTCTCGCC AAGGAACGAA GATAGTCTTG AATATCCAAG GTAAGTGAAA AAGAAGATGA	1740
TAGTCGGATA TGTTGGTATC CTCTATGTAG CACTTTAAAA GACATATTAT TCACCTATAA	1800
GGCTATCTAA TTGCTCTTCT TTTTCTCTT GCTCGTACAA ATATTTGGCA TCTTGCAAGA	1860
CATCGTCTCC ATGTTGCTTC ACAACAGAAA CAGATGCATC TAGCTCGTCT TTCAACTTGT	1920
AAGCCTTAGC CAAAGCTTTA GAATAACCTT TTTAGCTTC CTTACTTGCT AAGATTTTCA	1980



1105

AACCAAGGGT ACCAAATGCG ACACCACCCA AAAATAATGA AGATTTTTC GCAACTTTTG 2040  
CAACGGTTAA TACTTCTTTT AACATAGGG 2069

(2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4597 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAATCTTGCG CAATAAGCT CATCTCCATC TCCCGATTGA AACAGTCACT CCCC GGACTG 60  
TTTCAACGTC CCAAGACATA ATCTTAGGCA GATTCTCTAAA ATTACACTCA AAGTGGGAAGT 120  
CATTGAGCTT TCGAATGACA GTTGAAGTTG AAATGGCCAG CTGATGGGCA ATATCGGTCA 180  
TAGAAATCTT TTCAATTAAC TTTTGCGCAA TCTTTTGGTT GATAATACGA GGAATTTGGT 240  
GATTTTCTT GACGATAGAA GTTTCAGCGA CCATCATTTT CAAGCAATGA TAGCACTTAA 300  
AACGACGTTT TCTAAGGAGA ATTCTAGTAG GCATACCAGT CGTTTCAAGG TAAGGAATTT 360  
TATAGGGTCT TTAATGTCTA GTAATTTTGT GATAAAATGT AATTGTTCCA TATGATTCTT 420  
TCTAATGAGT TGTMTGTGCG CTTTTCATTA TAGATCTTAT GGGACTTTT TTCTACCCAA 480  
AATAGGCTCC ATAATATCCA TAGGGAATTT ACCCACTACA AATATTATAG AGCCCAAAGT 540  
TTTAGGTGCG TTGATAATAT GCGTTTMTTG AATTTTATAG ACTGCTCGTT TAAACTCTAT 600  
TTACTTCGTA CCTTCTGGAG CGAGACGGAA TATTAGTCAC ATACAAAATG AGTACTATTA 660  
GGATTTTATT TTCATGTACA ATTTCAGCCA GTCTTGTTAT AATCAGCCTA TAGGAATCAA 720  
GGAGGTGACT CTTATGGCTG TTTTGTGTC TTGGATGGA ATTGTGGTAG AAGTCCTTGA 780  
TGTCTTTTCT TCTTTTAATG GGGATAGTGA GTTTTCTTG TGTATAGCAT TTTGAATCTG 840  
GAATAGGACG CCATGACTGC TAAAAGATTT CTATAAATTA ATTTGATTTT CCTAATCAAT 900  
TTGTTTCATAT CTTATTTTCT TCCACTATAA ACGTCTTAAA GACAAGAGTC AGTTTGTAT 960  
GGAACGCTCT CAGTTCGAGG AGATGTTCCA ACTTCAAAGT AGTCGCTTGA CGACGCAAGA 1020  
AAAATTACAA TTGTTTACCT CTGTGTTTGC TGGCCGTTAT GATGTTTATG CTAAGAATTT 1080  
TATCAATGAA CAAGGGAAAA TTCAGTATTT TCCTTCCTAT GATTATGGTT GGAAGCAGTT 1140  
GCCACCTGAA AAACGGAGTT TCCAGACATT GACGAACTCC GTTTTGAAAT CTCATTTTTCG 1200  
TGGGGAGGCA GCTATCGGTA TCTTTCCTAT GCACTTAGAT GATAGCTGTT ATTTTGTGGT 1260

1106

ACTGGATTTG GATGAAGGAG ATTGGAAAGA AGCTGGTTTA ACCATTTCGAA GAATAGCCAG	1320
GGAACGCCAG ATGGAAGCCC ATTTAGAGAT TTCTCGTTTCG GGTACCGGAC TCCATATTTG	1380
GTTCTTCTTT GAGGAAGCGA TTCCGAGTCG AGAGGCTCGC TTGTTTGGAA AGAAACTGAT	1440
AGAACTGGCA ATGCAGGAAA GTATGCAACT GTCCTTTGAT TCTTTTGATC GCATGTTTCC	1500
AAATCAGGAT GTCCTTCCTA AGGGGGGATT TGGAAATTTG ATTGCCTTGC CTTTTCAAGG	1560
AGAAGCTTAC CATCAAGGGC GAACGGTCTT TGTGGATGAA CAGTTTCAGC CTTATGAAGA	1620
CCAATGGAGG TATCTACAAG AAATTCAGAG GATTTCAACT GCTAAAGTGG CACTGTTAAT	1680
CCAAGAGGAG TTAGGCAAGC AAGAATTGGA TAAGGAGTTG AAGGTCGTTT TATCCAATAT	1740
GATCCAACCTT GAAAAATCGT CTGTGACATC CAAGGCACTT TTTTCTTGAA AAATATGGCT	1800
TCCTTTTCTA ATCCCGAATT TTATAGTAGA TTGAAACTAG AATAGTACAC CTCTGCTTCT	1860
AAAACATTGT TAGAAATCGA TTTGACTTTC CTGATCGATT TGTCTGTTA TTATTTCAAT	1920
TTACTATATT TAAAGCAGGC TATGCGACAG CCAACCTATC AAATTCCTGA GAGAATGTAT	1980
TTATTTGGAG AATCCGATCA TTATTTATGG TTGCAAGAG GTTTGCTGTA TCCATTGCAA	2040
GATAAATTTA AGCAGGTATC TGTGGAAGAT AGGAGAAAAGG TACAAAGGTC TATTAGCGTG	2100
GAATTTAAGG GAGAACTCAC TTTTGAGCAA GAGTTAGCCC TGCAGATAT GACTTCTAAA	2160
GAAATGGTT TACTTCATGC GGAGACTGGT TTTGGGAAGA CCGTTTTAGG TGCTGCTCTT	2220
ATCTCTGAAC GGAAAACAAA AACAATTATT CTAGTCCATA ATAAGCAACT CTTAGACCAA	2280
TGGCTAGATC GCTTAAACTG CTTTTTGACT TTCGAAGAGG AGGAGGCTAT CCGTTATACG	2340
GCATCAGGTC GTGAAAAGGT AATCGGCTAT GTTGGGCAGT ACGGTGGGAC TAAGAAATGG	2400
CTGAGTAAAC TGGTTGATGT CGTTATGATT CAATCTCTAT TTAAGTTGGA AAATAGTCAA	2460
AGTCTTTTGG ATGAGTATGA GATGATGATT GTGGATGAGT GTCATCATGT CTCTGCCTTG	2520
ATGTTTGAAA AAGTTGTTGC TCAGTTTAGA GGAAGTATC TTTACGGTTT GACGGCTACG	2580
CCTGAGCGTA AGAATGGTCA TGAGCCTATT GTTTTTCAGA GAATTGGTGA GATACTCCAT	2640
ACTGCTGATA AGAGGGAAAC GGATTTTAAA CGGCAATTGC AATTAAGATT CACTTCTTTT	2700
GGTCATTTGG AAATTGAAAA GACCAAAGCA AGTAATTTTA TACAGCTTAG TGATTGGATT	2760
GCTACTGACT CAGTGAGGAA TCAGATGATT CTCAAGGATA TTCTAGCCCA AGTGGCAGAA	2820
GGACGGAATA TCTTGGTTTT AGTTAATCGA ATTCAACAGA TAGATGTCTT TGAAAAATTA	2880
TTGAAAGAGA AAGAGGTTGA TGA CTGTGTTAC ATTATTAGCG GAAAAACCAA AGTCCGAGAG	2940
AGAACGAGTT TACTGGAGAC GTTAGAACAG TTAGATAAAG GGTGTGTTTT GTTGTCTACT	3000
GGAAAATACA TTGGCGAAGG TTTTGACTTA CCTCAGTTGG ACACGCTTAT CTTGGCAGCA	3060

1107

CCCTTTTCTT	GGAAAAATAA	TTTGATTGAG	TATGCAGGTC	GGATTCATAG	AAACTACAAG	3120
GATAAGTCTT	TGGTGCGTAT	TTTCGATTAT	GTGGATATTC	ATGTTCCCTTA	TTTAGAAAAG	3180
ATGTTTCAGA	AACGACAAGT	AGCTTATCGA	AAGATGGATT	ATCGTGTCAT	CGAGGGTGAG	3240
GAGAAACAAT	TCGTTTATGT	TGATAGTAGA	TATGAGAAGG	TGTTGAGAGA	GGACTTAGCA	3300
GGGGAAAGAC	AGGAATGTCT	GCTTATTTTA	CCTTATGTGC	ACCAGACAAA	ACTGATGAAT	3360
TTTCTAAAAG	AATTTAGGAT	TAGTCAAATT	GAGATATGTA	TACCAGAGAC	GGTTGCAAAT	3420
AAAGCATGGC	TAGACCAGTT	GAAGAGCCAG	AAAATTAAAG	TGTCTTTTAC	TCAATCAAAA	3480
ATAGTAACGC	CTATTCTTTT	GGTGAATAAG	ACTATTGTTT	GGTATGGTGC	AATGCCATTA	3540
TTAGGGAAGG	TAGATGAGAT	GACCATATTA	CGTTTGGAAT	CAGCTAGTAT	AGTTTCTGAA	3600
CTAGTGGCAG	GTTTACGATA	GAGAAAAATT	TTAAAAATTT	CTATGTATGA	TTTTCATTTT	3660
TTTAGTGAGA	CTGTTGCCAT	TATCACATTC	GAATCACACA	AAATAAAAAA	ATTTTATATA	3720
GTACTTGACA	AATAGATTGA	AATATCATAA	AATAAAAACG	GTTACAGAGT	TATTAATTAT	3780
TTAAGCTTCA	TGTCACCATT	AAAAATTGAA	ATAAAAGGAT	GTTATCACTA	ATACAAGTGA	3840
GCAGGAACCT	ATTTAATCAC	ATCAGAAGAA	GTTTCTTGAT	GTTTTTAAGT	AGGTTCCCTT	3900
TATTTTAAAA	GGGAAATTTT	ATGATCATAA	AACGAATACT	AAACCACAAT	GCCGTAATTG	3960
CGCAAAGTAA	AAAAGATATC	GATATTCTTC	TTTTTGGAAG	GGGAATAGCT	TTTGAAGAA	4020
AAACTGGAGA	TAAAGTAAAT	CCAATTGATA	TTGAGAAAAG	TTTTTTTCTC	AAAAATAGAG	4080
ATAATATGAC	CCGTTTTTACA	GAGATGTTTA	TTAACGTTCC	TTTGGAGTTG	GTGTACATCA	4140
CCGAAAAAAT	AATTAACCTA	GGTAAAATAA	CATTGGGTAA	TAATTTTGAT	GAAATTATCT	4200
ATATTAATTT	AACGGATCAT	ATTTCTTCGA	GCATAGAACG	TTATAAAGAA	GGGATTATTA	4260
TTTCGAATCC	CCTACGCTGG	GAAATATCGA	AATATTATAA	AGAAGAATTT	GAACTTGGGA	4320
AAAGGGCTTT	ACAAATAATA	AAAAAAGAGT	TAGGTATTGA	ACTTCCAATT	GACGAAGCTG	4380
CATTCATAGC	GCTACATTTT	GTAAATGCTA	ATTTAGAAAA	TAATTTTCAA	GAGTCGTATA	4440
AAATCACTGA	AATAATTATG	GGAATTGAGA	AAATCATTCA	AGATTCTTAT	TGTACTGAGT	4500
TTAACCAAGA	TTCTATTGAT	TATTATAGAT	TCATAACTCA	TATGAAATTA	TTTGCCCATC	4560
GCTTGGTTGA	GAATACAAC	TATTGTGACG	ATGATGA			4597

(2) INFORMATION FOR SEQ ID NO: 176:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3984 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

1108

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

CGGCTTATTT ACTACTTGTT CCATCATATA TGAATATGC ATGAAACCTG CTCTCATATT	60
AGGGAATTTT TTATCCACTA AATAAGAGC TTGGTACATC AATGATTGC AAACAAAGGT	120
TCCTGCACTA TTGGATACAA CTGCCGGAAG TCCCTGTTTT TTGATAGCTT GTACCATCGC	180
TTTGATAGGT AAACACTAA AATAGGCCGA TGCTCCATCA ATACGAATCG GTGTATCAAT	240
TGGTTGATTG CCTTCGTTAT CAGGTATCGG AGCATCATCT TGATTAATAG CCACTCGTTC	300
AGGTGTTAAG CCGGTCTGCG CGCTGCTTG TCCAATACAA AGTACAGCAT CTGGTTGATA	360
TCGTAATATT TCTGCCTCTA AAACCTCTGA CGACTTATAA AAAACCGTTG GAATTTCTAC	420
CCAGCGAACT TCAGCCCCAT TAATCTCAGA TGGTAATAAT TTTACAGCCT CCAAAGCTGG	480
ATTAATCTTT TCACCTCCAA AAGGATTAAA ACCTGTAACC AATATTTTCA TTTTATTTTC	540
CTTTACTAAA ATGCGAGAAA GTACATTAAG AATATGTGAA TAACAATCAT TACTAGAGCA	600
ACACCTGCTT GAGCCTTTAT AACGCCATTC TGATCTTTCA TATCCATCAA TGCTGCTGGT	660
AGAGCGTTAA AATTAGCAGC CATTGGGGTC AATAAGGTCC CACAATAACC TGCTGTCATG	720
GCAAGAGCAC CAGCCACAAT TGGATTAGCT CCCAGAGCAA ATACAAAGGG AACTCCAACA	780
CCTGCTGTAA TAACGGTGAA TGCTGCAAAA GCATTTCCCA TAATCATTGT GAATAGAACC	840
ATTCCAAGAA CATAGGCCAA AACTCCTATA AAGCGACTAT CTGAAGGAAC AATACCCTA	900
ATCAGATGAG AGATAACATC ACCAACACCT GCTACAGTAA AAATAGCCCC CAAAGCCCCT	960
AATAATTGAG GAACAATCCC ACTTGTGAA ACTTGCTGAG TCATTGATT ATTTTCTGAT	1020
AACAGACTCT TAGGGTGACT ATTGTAATC ACAAGAACAG AAATTGTAGC AAACAAGGCG	1080
GCAAGGCTAA TCGAAATCTT GCTAAATTCT GGAATCATTT GCGCTAAGAC CAACGCAAGT	1140
ATTGCCATCA GCATAACTGG AATAAAAAAT TTATTTTCA ACCTGTTAGA TTCAATATTG	1200
GCTTTCATTT CATCTAAGGA TGGCAAGGTT CCGATACGGA CTGCTTAAA CAATGTTAAC	1260
AGCGATAATA GGATTACAAT AATACCAATA CTCATATTTG GCATATAGGA ACCACCTATA	1320
AACGTAATAG ACAATAGAGT CCAAAATGCA GATGTCCCAA GTCGAACTGG GTTTGTTTTA	1380
TCTTTATAAC TACAATAGGC TGTATGGAGA AATTGACAAC CAATCACAAT ATAGGTCAAC	1440
TCTAATAGTT GCTTTGCCAA CTCTGTCATT TTTGTTCTCC TCCCCTAGTC TTTTGTGATA	1500
TCAATTTTTT ATCAAATAAA TAATTATAAA TCCCCACTAC AATAAGTGTT ATAACAGCAA	1560
CAATAATAGA TGTAGAAGCA ATCCCTGCAT AATTGCTTTC ATAGCCTAAC TGATCTAATG	1620

1109

TTCCCCCTAT CAAGAGGACT CCCCCAGCAC CTACAAACGT ATTTTGAGCA AAGAAATTC	1680
CAAAATTTTC ATTGCGAGCC GCACGCGCTT TTATTGTCTC ATCTTCAACC TCTGTTAACT	1740
TTCTACCTAA TTGAGACTCT GCAGCTGCTT CTCCCATAGG TTGAACCAAA GGTCTGACAA	1800
ACTGAGGGTG TCCTCCTAGA CGAATTGAAA AGAAACCAGC TAACTCTCGA ATAAAGAAAT	1860
AAACTGTATA GAAGTTTCCA ACTGTCAGAC CTTTAATCTT TCGAATCAAA TCGATTGATC	1920
GTTGCTTGAG TCCAAAGGTT TCTGACAGCC CCACAAGAGG CAAGGTAACC ATAAAAATCG	1980
TGAGCACTCG CTGATTGCTA AATTCTTTTC CCAAAATCTC CAAAATTC ACGAGAGAAA	2040
CACCTGAAAC TAAAGCTGTA ACCAAACCAG CTAAGACTAC TGTGCAATT GTATCAAAT	2100
TTAAAATAAA ACCCACAACA ATGATTGCTA TTCCTATTAA TCTAATCCAC TCCATATCAA	2160
ACTCCTTTAT ATTCAAATG ACAGTATTTT TAAATTTTA TCAAGATCAA TACCATTCTT	2220
TATTTAATGT GTTTTCTAG TTCTTTTGG TATTGCTAT TGGATTCCAA TTTTCTTTT	2280
TGCCATTTT TAAAAACCTC GTTATATTCT TTTGTTGTAA CAATATCTTT TTGCAATTC	2340
ATTCCTTTAA AGATATATGG ATCCCCCTTA ATACCAACTT GTGAGTATGG TTTTGAGAAT	2400
GGTACTACGT TACTTACAAC TGGAGAACCA CCAGATGAAG CTGTTGGCAT CAATAATGAA	2460
CTATCTGTCG ACCAAGCTTG AGCTTTGGCA TATTTTTCAT ATCTTTTCTC TAGGTCAGTG	2520
GTCTCAGAAA CAGCATCTTC TAACAATTC TTATATTTAT CCAAACCAGG TTTAGTACA	2580
ACATCCTTAT CTTTTCCTTT CGTAATACCA AGGTGTTTCA TGGCAGAACC AGATTTTGGA	2640
TCTATAATAT TCAAGTGAGA CGCTGGATCT TGATAGCTTG GAGCCCATCC TGTACTGTTT	2700
AAATCATAGT CTTTGTGAGA AGGAGCAACA TTGCCGTATT TATCATTTTC CATCAAACCA	2760
TCAATAACAT TTCCAATAAC GTCTGTCCTC GATGTTGAG TCGCTATACT GTAGCCCAAT	2820
GATGCTGGAT CTAATGCATA GACATAAGAA AATGTTGTCG GTGCATCTGC TTCTTTATCA	2880
GTTTTTCCAC AAGCCACTAA AATAGCTGAC GTGCTCAGGA CCACTCCTGC TGTAAAGAGC	2940
CACTTTTTCT ATTTCAATAA GAATCTCCTT TGGTTTATTT TAATCTACTT TTACAATCCA	3000
ACCTTCTGGC GCTTCAATAT CGCCAACTG AATACCGTC AATTCATTAT ATAATTTACG	3060
CGTCACAGGA CCTACTTCTG TTTCACTATA GAATACATGG AAATCATCAC CATGTTGAAT	3120
ACCTCCAATT GGAGAAATAA CCGCTGCTGT ACCACAGGCA CCTGCCTCTA CAAAACGGTC	3180
AAGATTATCA ATTGGAACAT CACCCTCAAT AGGAGTTAAT CCAAGCGAT GTTCTGCCAA	3240
ATAAGCAAG GAATACTTGG TAATAGATGG CAAGATAGAT GGACTCAATG GTGTTACAAA	3300
TTCATTATCA GCTGTAATTC CAAAGAAGTT AGCTGATCCG ACTTCTTCAA TCTTTGTATG	3360

1110

AGTTGATGGG TCCAGATAGA TAACATCTGA GAAATGACGT GACTTGGCCA TTTTCCTGG	3420
TAAGAGACTT GCAGCATAGT TTCCACCAAC CTTAGCCGCA CCTGTACCAT TTGGTGCTGC	3480
ACGGTCGTAC TCATCCTGAA TCAAGAAGTT GGTGGGACC AAACCACCTT TAAAGTAATT	3540
TCCAACCTGGC ATAGCAAAGA TGGTGAAAAT GTACTCTTCT GCCGGTTTTA CCCCATAAT	3600
ATCTCCGACA CCAATCAAAA GAGGGCGAAG ATATAAGGTT CCACCTGTTT CGTATGGTGG	3660
TACGTATTCT TCATTGCGAC GGACAACCTGC TTTACAAGCT TCTACAAACA TGTCTGTCGG	3720
AACCTGTGGC ATCAAGAGAC GGTACATGT ACGTTGCAGA CGTTTAGCAT TTTCATCAGG	3780
ACGGAACAGT TGAACACTGC CATCCTTAGT ACGATAAGCT TTCAAACCTT CAAATGCTTG	3840
TTGTCCATAG TGAAGACTTG GAGAAGACTC TGAAATATGC AAAGTTGCAT CCTCTGTAAG	3900
CTCTCCTTGA TCCCATGTGC CATTTTGTGA ATGAGCAAGA TAGCGATAAG GTAATTTCAT	3960
ATAGGAAAAA CCGAGGTTTT CCGG	3984

## (2) INFORMATION FOR SEQ ID NO: 177:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 8703 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

TATCTAATTA TTGGTTTTTA TCGCTGACCT TGGCTATTGT TGGGGTTGTT TTACCCTTGT	60
TGCCTACAAC ACCTTTCCTT TTGTTGTCTA TTGCTTGTTC CTCCAGAAGT TCCAAGCGAT	120
TCCAAGATTG GCTTTATCAT ACCAAGCTCT ATCAAGCATA TGTAGCTGAT TTTCGTGAGA	180
CCAAGTCTAT TCGCGGTGAA CGAAAGAAAA AAATCATCGT CTCTATCTAC GTCTTGATGG	240
GAATTTCTAT TTATTTTGCA CCTCTTTTAC CAGTCAAAAT CGGTCTGGGT GCTTTGACCA	300
TCTTTATTAC TTATTATCTC TTCAAGGTCA TTCCAGACAA AGAATAGTTA AAACAGTAGT	360
TATTTGCCTT GATAAAATTG AAAGCATATT CATAACAATA TGATATAATA AAATTGAAGT	420
AATATTCAAG GAGAATCAAA TGATTTACGA ATTTTGTGCT GAAAAATGTA CTTTACTTGA	480
AAAAGCGATG CAGGCTGGAG CTCGTCGGAT TGAACCTGCT GATAATCTAG CAGTTGGTGG	540
GACAACACCC AGCTATGGAG TGAATAAGGC AGCGGTTGAA CTGGCAGCTA ACTACGATAC	600
AACCATCATG ACCATGATTC GGCCACGTGG TGGTGACTTT GTCTATAATG ACCTAGAAAT	660
TGCTATCATG CTAGAAGACA TTCGTTTGAC TGCTCAGGCT GGAAGTCAAG GGGTTGTATT	720
TGGAGCTTTA ACTGCTGATA AAAAGTTGGA TAAGCCTAAT CTGGAAAAGT TAATTGCTGC	780

1111

ATCAAAAGGA ATGGAAATTG TCTTTCACAT GGCCTTTGAT GAACTAAGTG ATGAAGATCA	840
AGCGGAAGCT ATTGACTGGC TCAGTCAAGC CCGTGTCACT CGTATCCTAA CTCGTGCTGG	900
TGTGTCTGGC GACTCCTTAG AAAAACGTTT TGTTCACTAT CACAGAATTT TGGAGTACGC	960
TAAAGGTAAA ATTGAAATTC TACCAGGTGG GGGGATTGAC CTTGAAAACC GTCAAACCTT	1020
TATCGACCAG GTGGGGGTAA CACAATTGCA TGGTACTAAG GTTGTTTTTT AAAAAATAGA	1080
AAGGAACTGC TAGCTTTGGG TAGCAGTTTT CACTTATGTT TGAAATTTTT AAATCCTATC	1140
AATTTAATCA AGAAAAGGCT CATGATTATG GTTTTATAGA AAATAGCGAA GTCTGGACAT	1200
ATAGTTGCCA GATTTTGCAA GGTGACTTTG TCATGACTGT GTCCATCACT GCTGATAATG	1260
TGAACTTTCA AGTCTTTGAC CAAGAGACTG GTGACCTCTA TCCTCACGTT TATATGAAAA	1320
GCATGAGGGG AAGTTTGTG GGAATGTCC GTGAGGCTTG TCTGGAGATT CTTTACCAGA	1380
TTCGGAAGGC TTGTTTGTG GTGCAAGATT TTATCTGTCA TCAGACTAAG CGTATCATGA	1440
CTCAAGTTCA GGAAAAGTAT GGAAACCACT TGGAGTATCT GTGGGAAAAA TCGCCTGATA	1500
CAGCTGTATT GCGCCATGAA GGCAATCAAA AGTGGTATGC CGTCTTGATG AAAATCTCTT	1560
GGAAATAAGCT GGAAAAGGGC AGAGAAGGAC AAGTGAAGC AGTCAACCTC AAGCATGACC	1620
AAGTAGCTAA TTTGCTTTCA CAAAAGGGGA TTTATCCAGC CTTCCATATG AGCAAGCGCT	1680
ACTGGATTAG TGTGTCCCTT GATGATACTT TATCAGATGA AGAAGTACTG GAATTGATAG	1740
AAAAAAGTTG GAACTTAACC TCTAAAAAT GAAATATTTT AATAATTTTC ATGAACTTTC	1800
AATTAGCTAA ATATTCTTTA CTGAAGAGAT TTTTAGAAAA TATAGGATTT ACCACACTAG	1860
AGGAATATGG TGCCATCTTC AAATACCTGA TTGAGAATGT CAAGACGGAT CGTCAGATCA	1920
TCTATTCGCC TCACTGTCAT GATGACCTCG GAATGGCAGT GGCAAATAGC CTTGCTGCTG	1980
TCAAGAATGG TGCAGGACGT GTTGAAGGGA CTATCAATGG TATTAGGGAG CGAGCTGAAA	2040
ATGCTGCTTT GGAAGAAATT GCAGTGGCTC TCAATATTCG CCAAGATTAC TACCAAGTAG	2100
AAACCAGTAT TGTCTTAAAT GAGACCATCA ATACGTCAGA AATGGTTTCT CGCTTCTCTG	2160
GTATTCAGT TCCTAAAAAC AAAGCCGTCG TTGGTGGCAA TACCTTCTCC CACGAATCTG	2220
GTATTCACCA AGATGGAGTC CTTAAAAATC CTCTCACTTA TGAGATCATC ACACCTGAAT	2280
TGTTGGTGT TAAGATTCTG CTTGGAAAAT TATCTGGTCG CCATGCTTTT GTTGAGAAAC	2340
TGAGAGAATT GGCCTTAGAT TTTACAGAAG AGGATATCAA ACCACTCTTT GCTAAGTTCA	2400
AGGCACTGGT CGATAAGAAG CAAGAAATCA CAGATGCAGA TATTCGAGCT TTGGTAGCTG	2460
GAACCATGGT TGAAAATCCA GAAGGCTTCC ACTTTGATGA TTTACAACCT CAAACTCATG	2520

1112

CAGATAATGA CATTGAAGCG CTCGTTAGCC TAGCCAAATAT GGATGGTGAG AAAGTCCAAT	2580
TTAATGCGAC AGGGCAAGGT TCCGTTGAAG CAATCTTTAA TGCTATCGAT AAGTTCTTTA	2640
ACCAATCTGT TCGTTTGGTG TCCTACACTA TCGATGCGGT AACAGATGGA ATCGATACCC	2700
AGGATCGGGT TTTGGTCACT GTTGA AAAACA GAGATACAGA AACCATCTTT AATGCAGCAG	2760
GGCTTGATTT TGATGTGTTG AAGGCTTCTG CTATTGTCTA TATAAACGCT AATACCTTTG	2820
TTCAAAAAGA GAATGCAGGT GAGATGGGAC GCAGTGTTC TTACCACGAT ATGCCTAGTG	2880
TGTAAAGGAG AAGGCTATGG CAAAGAAAAT AGTAGCTCTA GCAGGAGACG GAATTGGCCC	2940
AGAAATCATG GAGGTTGGTT TAGAAGTTCT GGAGGCTCTA GCTGAAAAA CAGGTTTTGA	3000
CTATGAGATT GACAGACGAC CGTTCGGAGG TGCAGATATT GATGCAGCAT GACCTCCCTT	3060
ACCTGATGAA ACCCTTAAGG CAAGTAGGGA AGCAGATGCT ATCCTACTAG TAGCTATCGG	3120
TAGTCCTCAG TATGATGGAG CAGTGGTTCG CCCTGAACAA GGCCTGATGG CTCTCCGTAA	3180
GGAATCAAT CTTTACGCTA ATATTCTGCC TGTAAAAATC TTGACAGTC TCAAGCATTT	3240
GTCACCACTC AAAGTGAAC GAATTGCTGG TGTAGACTTT GTCGTGGTGC GTGAATTGAC	3300
AGGCGGGATT TACTTTGGAT ATCATATTCT TGAAGAGCGC AATGCGCGTG ATATCAACGA	3360
CTATAGCTAT GAGGAAGTGG AGCGGATTAT TCGCAAAGCC TTGAAATTG CAAGAAATCG	3420
CAGAAAAATC GTTACTAGTA TCGATAAGCA AAATGTTCTA GCGACCTCAA AACTCTGGCG	3480
GAAAGTAGCT GAGGAAGTCG CACAGGATTT CCCAGATGTA ACCTTGAAC ATCAGCTGGT	3540
AGACTCAGCT GCTATGCTTA TGATTACCAA TCCTGCTAAG TTTGATGTTA TTGTAACGGA	3600
GAATCTTTTT GGAGATATTT TATCTGATGA ATCAAGCGTC TTATCTGGTA CACTTGGGGT	3660
TATGCCATCA GCCAGTCATT CTGAAAATGG ACCAAGTCTC TATGAACCTA TTCACGGTTC	3720
AGCACCTGAT ATTGCAGGTC AAGGAATTGC CAATCCTATT TCCATGATTT TATCAGTTTC	3780
CATGATGTTG AGAGATAGTT TCGGACGTTA TGAGGATGCA GAGCGTATCA AACGTGCTGT	3840
TGAGACAAGT CTGGCGGCAG GAATTTTAAC GAGAGATATA GGAGGTCAGG CTTCAACAAA	3900
GGAAATGACG GAAGCTATTA TTGCAAGGTT ATGAAGTTAG ACGAAAAAAT TACTCTAGTC	3960
CTTTTGATTT GGAATGTCAT CATTTTCTTG ATTTATGGTA TTGACAAATC TAAGGCAAGG	4020
AGAAGAGTTT GGCGCATCCC TGAGAAAAATC TTACTTATTT TAGCCTTTAC TTTTGGTGGT	4080
TTTGGTGCCT GGCTAGCAGG AATCATCTTT CACCACAAGA CTCGAAAATG GTACTTTAAA	4140
ATAGTTTGGT TTCTTGGGAT GGTGACCACA CTAGTAGCCT TATATTTTAT TTGGAGGTAA	4200
TGGATGGCAG GGTCTTCGAG GGAATACGCT GCTTGGGCTC TAGCGGACTA TGGTTTAAAG	4260
GTCGTGATTG CAGGATCTTT CGGTGACATT CATTACAATA ATGAACTCAA TAATGGCATG	4320



1113

TTGCCAATCG TTCAGCCTAG AGAGGTTAGA GAGAACTAG CCCAGCTAAA ACCAACCGAC	4380
CAGGTAAC TG GACTTGGA ACAACAAAA ATCATCTCAC CAGTTGAAGA ATTCACCTTC	4440
GAGATAGATA GCGAGTGGA ACATAAACTC CTAAATAGTT TGGATGATAT CCGTATTACC	4500
TTGCAGTATG AAGAGTTGAT TGCTGCTTAT GAAAAACAAC GACCAGCCTA CTGGCAGGAT	4560
TAGAAAAAT AGAAAAGGAG ATATAGTAAA CTGAAATAAG ATGTAAACAA ATGAATTGGA	4620
GCTTAACATC CATTTCAGC AATTTTITAG AACTACAGT GGACTATTCT GGATTCAACA	4680
CATTATAAAA TTATGACAAA ACACATTCAC AAGAAGGCTA CGACATTTTA AAAGGTGAGG	4740
GCGGATGTAT CGTTTGCCCT ACTAAAGTTG GTTACATTAT CATGACCAGT GACAAGGCAG	4800
GACTTGAGCG TAAGTTCGCA GCCAAAGAAC GTAAGCGTAA CAAACCAGGT GTTGTCTCT	4860
GCGGTAGCAT GGATGAACTT TGCGCTTTAG CGCAACTCAA CCCAGAAATT GAAGCATTC	4920
ACTAAAAACA TTGGGATGAA GATATCTTC TTGGTTGTAT CCTTCCTTGG AAACCAGAAG	4980
CCTTTGAAAA ACTCAAAGCA TACGGGGATG GCCGTGAAGA ACTTATTACT GATGTACGTG	5040
GTACTAGCTG TTTTGTATC AAGTTTGGA AAGCAGGTGA ACAATTGGCT GCCAAGCTTT	5100
GGGAAGAAGG TAAATGGTC TACGCCTCAT CTGCTTCAAT GACAAAACGA TTGAACTCG	5160
CTATGAGCAA GGTGTAATGG TGTCTATGGT CGATAAGGAC GGCAAACTCA TCCCAGAACA	5220
AGGAGGAGCA CGTTCAACTT CACCAGCTCC AGTTGTGATC CGTAAAGGGC TTGACATTGA	5280
TAAATCATG ATGCACCTGT CAGATACTTT TAACTCATGG GACTACCGTC AGGTTGAGTA	5340
TTATTAGGAT AGAGAAGAAG TCTAGTGTTA TGAGATATTA AAGCTCCTAA CACTGGGCTT	5400
TTGTTTAGAA TTTCTTTCT TTTCTATAG GATATGGTAT TCTATGTAGA AAATATATGT	5460
TAATAAGTAA TGCCAATATT TAAACATCAT TAGTAAAAGG AGTTAGATTG ATGAATAAAA	5520
GAAAAGTTAG TTTAGAAGAT TTTATAAAT GGTATAGTCT AAATAAGAA GAGTTATTAA	5580
ATAAGGCAAC TGTGGTGAA AAGTTTAATG ATAAATTAAA AGAAGAGTTT CTCCAGGAAT	5640
GGCCTTTGGA TAGGATTTTA ACAATGTCAA TCGATGAATA TGTAATAGGA AAGGGACAGC	5700
AAAATAAGTC TTTATGCTAC GCTCTTGAGA AGGGAAAATA CAAAATCTA TTTCTTGGA	5760
TTTCTGGTGG CTCAGCTTCA AAATTTGGTA TTTATTGGAA TAAAAACA AACAAATATA	5820
AAGATCAAGC TAATAATGAG ATTCAGAGT TGGATCAGCG ATTTTCAAAA TTAATCAG	5880
ATTTGTATGA AATTATCAA GAAGGTATTC GTTTTAACTT TGAAATCCT ATTTTGTATA	5940
TGAAAAGATC AACAAATGAA TTTATTGGTC GTTCTGCTAT GGTGACAAAA TTACTTTGTA	6000
TCTATACTGA GGGAGATCCT TTCTTTGGTG TAAATATTAA TAGTCAGAAA GAATTTTGGA	6060

1114

ACCACTTTGT TTCTCAGACA AATCAAGGTG GACCTTATCT GCAAAATCAT AAAATAATTG	6120
AACTGGTGTC CAAAACCTTAT CCTGAGTTGG AGCCATCGAA ATTAGGAACT ATGCTTTTTG	6180
AGTATTCTAA GCTTTTTATG GAAAATAAGG AAGACAATAG TACAATGGAT TCATCAAACA	6240
ATTTTCGTCA TCAATTAACCT CAATCTCTAT TAAAGTCTCC AAACCTCATC CTCCGCGGTG	6300
CTCTGGCAC GGGAAAAACT TATCTTGCTA AAGAAATTGC TAAAGAATTA ACGGATGGCA	6360
ACGAAGATCA AATCGGATTT GTACAATTC ACCCATCATA TGATTATACG GATTTTGTAG	6420
AAGGTTTAAG ACCAGTATCA AATGGGGATG GAGCTATTGA GTTTAGGCTA CAGGACGGTA	6480
TTTTTAAAGA TTTTGTCTAG AAAGCAAAAG AAACCCAATT GATTGGAGGA CAAGATAATT	6540
TTGATGAGGC TTGGGATTCT TACTTAGAAT ATATAAATGT TGCTGAAGAA AAAGAATATA	6600
TAACAAAAAC ATCTTACTTA TCTGTTAATA GTAGACAAAA TTTGTCAGTA AATTATGATA	6660
GTGGTGTTCC AGGATGGTCA CTACCTAGCA AATATGTTTA CGAGTTGTAT AAAGATAAAA	6720
ATTATAATAA GCAAGAATAC TACAAAAGTG GTGGAAAAAC TGTCCTAGAA ACATTGAGAA	6780
AGAGATTTGG TTTGAAAGAC TATGTTTCCC CAACAGAAAT TGATACTGAT AAGAATTTTG	6840
TCTTCATCAT CGATGAGATC AATCGTGGGG AGATTTCTAA GATTTTGGC GAACTCTTTT	6900
TCTCTATCGA CCCCCTAT CGTGGTGAAA AAGGAAGTGT TTCTACCCAA TATGCAAATC	6960
TACACGAAAC TGATGAAAAG TTCTATATCC CCGAAAATGT TTACATCATC GGAACATGA	7020
ATGATATTGA TCGTTCAGTG GATACCTTTG ATTTTGCTAT GCGTCGTCGT TTTCTTTTG	7080
TTGAAGTTAC TGTCGAGGGT CAAGCTGGCA TGTTGGATAA AGAGTTGAAT ATCCATGCG	7140
AAGAAGCAAA AATTCGTCTA AGAACTTGA ACGCTGCTAT CGAAAATATT CAGGAATTAA	7200
ACAGTCATTA TCATATTGGA CCAAGTTATT TTCTTAAGTT GAAGGATGTA GATTTTGACT	7260
ATGAATTACT CTGGTCTGAT TATATTAAGC CTCTCCTAGA AGACTACTTG CGAGGTTCTT	7320
ATGATGAGGT TGAACTTTG GAaACTTTGA AAAAAGCATT TGATCTGACA AATAATGAGC	7380
AAAAAGATCA GGCAGTAGCT GATGACAATG AAGGCGATGA AAACGATGAT GCGGATTAT	7440
GATAATCAAC ACAAGATTAT TAAAGAAAAA TTTGTTGAAG AATATCCTAA ACTAAGCAAT	7500
CCTCTTTTAG ACAGAACCTT GGAAAGTCTA TCCCAAGATG AACGTATTTT CATTTTCCA	7560
AATGATTWGA CTCATACTCC TGATTTGGAT AAGGACCAAA AGATTTTGA AACAGTCAAT	7620
CAGAAAATCA AGACAGGGAA CGTGATTGGT TTTCTTGAT ATGGTCAGGA AAGATTAACG	7680
ATTTCTCAC GATTTTCTGA TGAGAGTAAT GACCACCTTT TGCATTATCT CTAAACAAG	7740
GTCTTCATA TCAATCTCAC TAGTTTAGAT GTTGCTTTGT CTCGTGAAGA GAGGCTTTAT	7800
CAACTTTTGG TGATCTCTT TCCCAAGTAT CTACAAGCTG CTATTCGAAA AGGTCTTTAT	7860

1115

AAGGAATATC ATCGATTTTC TCATAACGAC AGTCATGTTA AGGGAGTGAT TGATGTAAGA	7920
AACCATCTCA AGAAAAATCT TCCTTTCACG GGAAATATTG CCTACGCAAC GAGAGAGTTC	7980
ACCTATGATA ATCCCCTCAT GCAGTTGGTC CGTCACACTA TTGAATACAT TAAGAATCAG	8040
AAAAGCATTG GTCAAGGGGT ACTAGATAAT CTCTCAACTA GTCGTGAAAA CGTATCTGAA	8100
ATCGTGCGTG TAACGCCCTC TTATAAACTA GCTGATCGTG CTAAGATTAT TCGGGGAAAT	8160
CAATCTAAAC CTATACGTCA TGCATACTTT CACGAGTACA GAAACTTACA AGAACTTTGT	8220
CTGATGATCC TAAACCAAGA AAAGCACGGT TTAGGGTATC AAGATCAAAA AATCTATGGT	8280
ATTCTCTTTG ATGTTGCCTG GCTTTGGGAA GAGTATGTTT ACACCTTGTT GCCAAAAGGT	8340
TTTGATACATC CCAGAAATAA GGATAAGACG GATGGAATTT CAGTATTTTC TGTGGGAAA	8400
CGAAAAGTAT ATCCAGATTT TTATGACAGA GAACGAAAGA TTGTTCTAGA TGCAAAATAT	8460
AAAAAACTGG AATTGACTGA AAAAGGAATC AACCGTGAGG ACTTATTCCA GCTGATTTC	8520
TATTCTTATA TTTTAAAAGC TGAGAAGGCT GGAAGTATTT TTCCTAGTAT GGAGCAGTCA	8580
GTAAATAGTG AAATAGGAAA AGTAGCTGGC TATGGAGCTC AATTGAAGAA GTGGTCTATT	8640
CGAATCCCTC AGAATGCCTC ATTCTATAGT ACATTTTGTA AAATGATGGA AAATTCAGAA	8700
GAG	8703

## (2) INFORMATION FOR SEQ ID NO: 178:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4854 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

CATCACCAGT TTTAGATGGC TTAAACAGTG AAATTATTGC TTTTAATCTT TCTTGTTTCG	60
CTAATTTAGA ACAAGTACAA ACAATGTTGG AACAGGCATT CAAAGAGAAG CACTACGAGA	120
ATACGATTCT CCATAGTGAC CAAGGCTGGC AATATCAACA CGATTCTTAT CATCGGTTCC	180
TAGAGAGTAA GGAATTCAA GCATCCATGT CACGTAAGGG CAACAGCCAA GACAACGGTA	240
GGATGGAATC TTTCTTTGGC ATTTTAAAAT CCGAAATGTT TTATGGCTAT GAGAAAACAT	300
TTAAATCACT TAACCAATTG GAACAAGCCA TTATAGACTA TATTGATTAT TACAACAATA	360
AGAAAATTAA GATAAACTA AAAGGACTTA GTCCTGTGCA GTACAGAACT AAATCCTTTG	420
GATAAATTAT TTGTCTAACT GTTTGGGGGC AGTACACAAG AAAGCGCTTT AAAACCAGTA	480

1116

GACCTTTTCA TAAGGTTTCGC TTGATGTACC AAGATGAGGC TGGTTTCGGT AGAATCAGTA	540
AACTGGGATC TTGTTGGTCT CCAATAGGAG TAGGTCCACA TGTCCATAGT CACTATATAC	600
GAGAATTTTCG CTATTGTTAT GGAGCTGTTG ATGCCCATAC AGGCGAATCA TTTTCTTAA	660
TAGCTGGTGG ATGTAATACT GAGTGGATGA ACGCCTTTT AGAAGAGCTT TCACAAGCTT	720
ATCCAGATGA TTATCTTTTA CTCGTTATGG ACAATGCTAT ATGGCATAAA TCAAGTACCT	780
TAAAGATTCC GACTAATATT GGTTTTACCT TTATTCCTCC ATACACACCA GAGATGAACC	840
CATTGAACAA GTGTGGAAAG AGATTCTGTA ACGTGGATTT AAGAATAAAG CCTTTCGAAC	900
TTTGGAAGAT GTCATGAATC AACTCCAAGA TGTACATAAA GGATTGGAGA AGGAGGTGAT	960
AAAGTCCATC GTTAATCGGA GATGGACTAG AATGCTTTTT GAAAACAGAT GAGTATAAAA	1020
TTGAATTGCT TATAAAAAAG CTCATACAC TGGATGTGTA TAGAGCAATG GGGCTTTATT	1080
TGATATAGAG TTCTTGGTTT TTTAGGACAA TTTCTCGGAT ACTTGCAAAC TTTTAAAGTT	1140
TTTTGATTTC TTCTGGATGA GTGACGAGAG TGATAACATA ACCTTCCTTG CCCATACGAC	1200
CAGTACGGCC AGCACGGTGT GTGTAGGTTT CGCTATCTCT AGGAATATCA AAGTTTACGA	1260
CACATTCTAG GCTATCGATA TCAATTCCAC GAGCCAAAAG GTCAGTTGCA AGAAGCAGGG	1320
TTAGTTGGTT ATCTTTAAAC TTTTCTAAGA TGATTTTCT AAATTTAACA TTAACATCAC	1380
TAGCGAGGGA AACAGCCAAT ATATCAGGAT ACTGTAGTTT TTCCTCGGCA TTCCCAAGGT	1440
CTGACAGGCT ATTGAAGAAG ACTAGACCAC GGAAATCCTC TACATGAGCC AGTTTTCGTA	1500
GCATATCCAC TCGATGACGT TGGTCTACCT GCATGTAGAA ATGCTGGATA TTGTCCAATT	1560
TTTGATCAGA GAGATCAATA GTGCGTGAT TCGGCACAAT CTTTCTTGG TCAAACTTGG	1620
TCGTGGCACT CATGTAGACC AGTTGGTGGT CACGAGGTGC GTAGTGAGTG ATTTTTTCTA	1680
CAAAGTGAAT CTGAGAATCA TCTAGTAATT GGTCAAATTC ATCCAGGATG ATGGTTTCCA	1740
CATTATCAT CTTGATTTTT TTAAGTTTAA TGAGTTCAA GATACGGCCA GGAGTTCCAA	1800
TCAGAATTC TGGCCCTTT TTAAGACGTT CAATTGTGCG TTTCTGACTT GAACCTGAAA	1860
GGAAGAGTTG AGCAGTCAAT CCGATAGCTT CTGCCACGT TTTACATAA TCAAAAATCT	1920
GTCCAGCAAG TTCCGTATTT GGTGCTAGAA TCAAGAGTTG TTGGGCTTTT TTCTTTTGTA	1980
GTCTGAGAAG ACTTGGTAGG AGATACGCTA GGGCTTACC AGTTCCGGTT TGGCTCACTC	2040
CTAGGAGGTT TTCTCCAGCA AGAAGGGGCT CAAATAGTTG AGTTTGAATG GGGGTGAATT	2100
CTTGAAACC GAGTTGGTCA CTCAGTTCTT GCCATTCACT CGGTAGTTTG GTTTTCATTT	2160
TTCTGCCTCA AATCTAATGC CAGCAGTCTG GCGCATGGTA TATAGTAGCT CATGAACAGA	2220
GCCTGCATCA TACAGCCAAG TTTGGTAGAG ATTCAGATCT GGTGCTGGA TCATGTGTGC	2280

1117

AAATGCAGCG ACTTCCTCAG TCATCGTATG AGGAGCCTGT TGGATAGGAA GCTGGACTTG	2340
ATTTCTTTGG TGGTCGGTAA AAATAGCTGA GCGAATATGC TCAATCGTGT TGAGAGTCAA	2400
GGTTCCATCT GTTGTATAAA TCTCGCAAGG AAGATTGGAA GTGATGTTTT TTCCAGCCTT	2460
GATGTGAACT TGATAGTCTG GGTAGAAGAG GATACCATCT CCATTTAGGT CAATGCTATT	2520
GTCAAGCTGT TGAGCATGGT AAGTCGCGTC ATTGGCTTTT CCAAAAAGAC GAACAGCAGC	2580
ATAGAGGGGA TAAATCCCCA AATCCATGAG GGCTCCACCA GCAAAACGGT CTGAAAAGAC	2640
ATTTGGTGTT TGTCCAGCCA ACAAGTCAGG CATCTTGGAA GAGTATTTGG CATAGTTGAA	2700
ATCTGCTCCT AACACTTGCT TATCTGCTAA AAAGTTTTTG ATAGTAGTAA AGGCTTTCTC	2760
GTGGTAATTA CGAGCTGCTT CAAAGATAAA ACAGTTATTT TTTTCAGCTG TTTGAATCAA	2820
ATCAAACCAT TCTTGTGGTT GAGAGACAGC TGGCTTTTCG AGAATAACAT GTTTACCAGC	2880
AGACAAGGCA GCTTTTGCCT GAGCAAAATG TAAGGAGTTT GGACTGGCGA TATAGACTAA	2940
ATCAAAAGAA GATTTGAAGA AGACTTCTAA TTGATCGAAT AGTTGGATAT TCTGATAGCG	3000
AGAAGCAAAG GTTGCTGCAG TTTCTAGTTT TCTAGAATAG ATTGCGACCA GTTGGTATTC	3060
TCCACTGGTA TGGGCTGCTT CTATGAAATG ATGGCTGATA GCGCCAGTTC CGATGACACC	3120
TAATTTTAGC ATAAATACTC CTTTTCCGAT TTAAATCCT TCTTTCATTA TAACATAGAT	3180
AGACGGGACT ATCCAACAGA GAGGAGAAAA TTTCAAATAA GCTATTAGCT TTCTTTTCCG	3240
AATAAATAGA TAGAAGCATA GAATCTAGCA AACCTAGATT TAAAAATGTG CTATAATAGA	3300
AGGAGGAAAA GGAGGATTCT CAGACATCTA GGTATCAGCC CAACTAATGA TTTGTCAATT	3360
TATCCGCGAT ATGCTGGACT TGCCAGCAAA AAATGTGACG ATTTTGGAGG GAAGTAACAT	3420
TCACGTCTTG CTTTCCATGC CCTACTCAGC GTAAGATTTT TATACTAGTA TAGACGTCTT	3480
GGCGGAGTTA GATAATGGAA TCCAAGTTAT CATCGAAATT CAGGTTTCATC ATCAGAATTT	3540
TTTCATCAAT CGCCTATGGC CTTATCTGTG CAGTCAGGTT AATCAAAACC TAGAAAAAAT	3600
TCGCCAACGT GAAGGTGATA CCCACCAGAG CTACAAACAA ATCGCACTAG TATACGCTAT	3660
CGCAATTGTC GATAGTAATT ACTTCTCAGA TGACCTAGCT TTTCATAGTT TTATAGTAAA	3720
ATGAAATGAG AACAGGACAA ATCGATCAGG ACAGTCAAAT CGATTTCTAA CAATGTTTTA	3780
GAAGTATAGG TCTACTATTC TAGCTTCAAT CTACTIONAGAA TTCCATAGAT AGAAAACTAC	3840
ATAATCTCTA CAGATACGGA TGTTGGAGTT GATGTAAGAT GCTTTGGCTT GCTAGAGGAA	3900
TTGTGGATTG CCAAATTGTA TCATTGAAAT TATTGCTCAA ATTTGTTATG ATATAAATAT	3960
GAATAAAAGT AGACTAGGAC GTGGCAGACA CGGGAAAACG AGACATGTAT TATTGGCTTT	4020

1118

GATTGGTATT TTAGCAATTT CTATTTGCCT ATTAGGCGCA TTTATTGCTT TTAAGATCTA	4080
CCAGCAAAAA AGTTTTGAGC AAAAGATTGA ATCGCTCAAA AAAGAGAAAG ATGATCAATT	4140
GAGTGAGGGA AATCAGAAGG AGCATTTCG TCAGGGGCAA GCCGAAGTGA TTGCCTATTA	4200
TCCTCTCCAA GGGGAGAAAG TGATTTCCTC TGTTAGGGAG CTGATAAATC AAGATGTTAA	4260
GGACAAGCTA GAAAGTAAGG ACAATCTTGT TTTCTACTAT ACAGAGCAAG AAGAGTCAGG	4320
TTTAAAGGGA GTCGTTAATC GTAATGTGAC CAAACAAATC TATGATTTAG TTGCTTTTAA	4380
GATTGAAGAG ACTGAAAAGA CCAGTCTAGG AAAGGTTTAC TTAACAGAAG ATGGGCAACC	4440
TTTTTACACTT GACCAACTGT TTTGAGATGC TAGTAAGGCT AAGGAACAGC TGATAAAAGA	4500
GTTGACCTCC TTCATAGAGG ATAAAAAAT AGAGCAAGAC CAGAGTGAGC AGATTCTAAA	4560
AAACTTCTCT GACCAAGACT TGTCTGCATG GAATTTTGAT TACAAGGATA GTCAGATTAT	4620
CCTTTATCCA AGTCCTGTGG TTGAAAATTT AGAAGAGATA GCCTTGCCAG TATCTGCTTT	4680
CTTTGATGTT ATCCAATCTT CGTACTTACT CGAAAAAGAT GCGGCCTTGT ACCAATCTTA	4740
CTTTGATAAG AAACATCAAA AAGTTGTCGC TCTAACCTTT GATGATGGTC CAAATCCAGC	4800
AACGACCCCG CAGGTATTAG AGACCCTAGC TAAATATGAT ATTACAAGCG GGGT	4854

(2) INFORMATION FOR SEQ ID NO: 179:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2186 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

TAAACAGGTG TTAGGTGCTC TAACTATTA AAATTCTAAG GAAATAAGGC TACTTTTCT	60
GGGTCTTGTT CATAGTAGGT GTGGTTCTTT TTTTCGAGTG TAGCCCATAG CTTTGAGCGC	120
ATAGTGGATG GTAGTTGGAT GACAGCCAAA TTCAGAAGCT ATTTGAGTCA AATAAGCATC	180
TGGATTGTCA GTAAGATAGT TTTTAAGTCT ATCTCTATCA ACTTTTCTTG GTTTTGTTCC	240
TTTTACTTGG TGGTTTAGCT CTCCTGTTTT CTCTTTTAGC TTTAACCAGC CATAAATGGT	300
ATTACGTGAG ATTTGAAAA CGTGTGATGC TTCTGTTATA CTACCTGTTT GCTCACAATA	360
AGAGAGAACT TTTTACGAA AATCTATTGA ATATGCCATA AGAAGATTAT ACCACATTGT	420
GTAATATTTT TGGTTTATTT TACTATATTT CTAAACACTT AGAAATAATA AAACAAATTA	480
AATATTATTT CTAAATATTT GAAAATAACA TCTATTTGTA TTATACTATC TTTGAGGTAA	540
CTATTATGAA CTATATCAAA AGACCACATT ATTTAGATTT TTTAAGAAAA CATCGTGACC	600

1119

GACCAATCAT CAAAGTTGTG AGTGGAGTTA GACGAGCTGG TAAATCTGTG CTTTTTCAAC	660
TCTATAAAGA GGAGTTACTA GCAACTGGGG TAGACGAGGA TCAGATTATA TTCATCAATT	720
TCGAAGATTT GAGTTACTAT GATCTGCGAC ATTTTCAAAC ATTATTGCT TATATAAAAG	780
ATCAATTAGT TAGCAAGAAA ACATACTATA TCTTTTTAGA TGAAATTCAA TATGTTGAAA	840
AATTTGAACT GGTAGCAGAT AGTCTATTCA TCTTAGCAAA TGTAGACCTC TATTTGACTG	900
GATCTAACGC CTACTTTATG AGTAGCCAAT TAGCAACAAA CTGACTGGT CGGTATGTTG	960
AGATAGAGGT TCTTCCTTTG TCATTTGAAG AATATCTATC AGGTCAATCT CTCACAGAGA	1020
ATCTGAATAC AACAGAAATT TTTAACAATT ATCTCTTTAG TGCTTTCCT TACTTATTGC	1080
AAACATCATC TTACGATGAA AAAATTGACT ATCTCAGAGG AATATATAAC TCCATACTGT	1140
TAAATGATAT TGCTACTAGA TTGGGAAAAC CAAATCCTAC TATTATTGAG CGCATTTGCC	1200
GAACCCCTCT CAGTAGTACA GGTAGCTTAA TATCAACAAA TAAGATTGCGC AATACCCTAG	1260
TCAGCCAAAA TGTTTCAATA TCCCATAATA CTTTGGAAAA TTATTTGACA ACTTTGACAG	1320
ATAGTTTACT TTTTATTCC GTTCCACGTT TTGATGTAAA AGGTAGAGCA TTATTGCAAC	1380
GTTTAGAAAA ATATTATCCC GTTGATTTAG GTTTACGACA TCTCTTATTA CCAGACCAGA	1440
AAGAAGACAT TAGGCATATC TTGGAAAATA TGGTATATTT GGAATTGAGA CGTAGATATT	1500
CACAAGTATA TGTTGGTAAT TTAGATAAGT ATGAGGTTGA TTTTGTGTT GTAACGTATC	1560
TTGGCCACTA CGCTTATTAT CAGGTCAGTG AAACAACACT TGCTCCAGAA AACTAGAAA	1620
GAGAACTTAG ACCACTAGAA GCCATTAAAG ATCAATTCCC TAAATATCTA TTAACAATGG	1680
ATACGATTCA GCCAACAGCC AATTACAATG GAATCGAGAA GAAAAGCATT ATAGATTGGT	1740
TACTAGAAAA ATAGATAAAT ATAAATCATA CAGCTAATTA GATTTGCAAC AGTCTGTTAT	1800
CAATGATTCT ACCCAAATCC TAACAAGATA TAGTGAATTT CGAATACGCT ATATAATACG	1860
GACACTTGAA AATAGAAATT GGGGATGAAA GGGGATCTAT AATTCTGGA AGTACTATCA	1920
AAAATTAATA TCATAGTCTT ATTAGAGAAT AGCATCACCC ACTTTCTCAA ATAAGATTAA	1980
ATTGTAAC TG AATTATAATG AAAAAGAGAC TGAGCAATCA GTCTTTAAAA TCAGAAAAGC	2040
GCATAGTATC AGGTATTGAA CAACCTTGAT AATATGCGTT TTATTATGGA AATATTTGCT	2100
TCATTTTCTC CTGAAATAGA GCTTTTGCTA TCCTATTTTT CTCTATTTCT AATGATTAC	2160
TTCAACTTCT TACCTCTTGG GAAAAA	2186

(2) INFORMATION FOR SEQ ID NO: 180:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3236 base pairs

1120

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

GTCACACGTT TGA	CTTCACG	TATTT	CATAA	GTATAA	ACTT	TATTTTT	TATC	GGTTAG	AATA	60
ATCTTCATGC CAT	TTTTTAGC	ATTATC	TAAA	GGAGAAA	ATA	ACATTTT	TATT	AGCATT	TATCA	120
ACACCAAAGA TAT	GGTGACT	AGCTAG	ACTA	TAATTT	CCTT	CTCCC	ATTAC	TTGCTC	GCGT	180
TTCATTGTAC CAG	CTCCGT	GAAGAG	ATTA	ACATTAT	CAA	GTCC	TTTAAA	AATCGG	CAAA	240
TTCATTTCCA ATT	CAGGAAT	TGCAAT	TCCC	CCAATA	ACTG	GTAATTT	TTTG	AGCATC	CCCAT	300
TGAGAAGTTA GA	CAGCTTC	CGAAG	AGATA	GCTTTG	ACAG	AATCAA	AGTC	AAAATT	GCCT	360
TCTGTATCCT GAT	TTTCTTC	TAATTTT	TCT	TTTGAT	ACCT	GGCTAA	CTTG	ATACTT	TATTG	420
GTATTCCAGA CT	ATGAAAAT	ATTT	CGAATT	TGAGTAT	TAA	AAATCAA	AGC	CAGTGAC	AGT	480
AATATCAGAA AT	CCTGCTAG	GATATTT	GTG	AGCAGAT	TTT	TTGCTT	GTG	TTTCTTT	TTTA	540
TTATTTTTTT GAG	ACATTAT	GCTTCAC	CTT	CTGTTT	CGTT	TTCTGT	CCCA	ACTTCT	TCTT	600
TTTCTGCCAC CG	CAACCGTT	GTGAA	AGTCA	CTATCT	GAGC	ATCTTG	ATCC	AGGCGC	ATTA	660
CTTTAACTCC CAT	AGTTGCA	CGTCCT	GTTT	GTGAA	ATATT	GGCAAG	ATTG	GTTCGA	ATCA	720
TGACACTGT AT	CAGTGATA	ATCAT	CAAAT	CCTCAT	CCCC	TTGAAC	AGTC	ATAAGAC	CGG	780
CCAGCAAGCC AT	TTTTTTTCG	GTAATTT	TAG	CTGTCT	GCAT	TCCCTT	TACCA	CCACGAC	CTT	840
TTGTTGGGTA TT	CAGTAGCG	ACTGT	ACGCT	TACCAT	ATCC	TTTTTCT	GTG	ATAATA	AGAA	900
CCTCATCTTG AT	CAGTAATC	AAGCTG	GCAC	CAACA	ACTGT	GTCTCCT	TCA	CGAAGG	TTAA	960
CACCTTTCAC ACC	AGTGGCG	ATACGG	GCTCA	TACCAC	GAAAC	GGCTGAT	TGA	TTAAAG	CGAA	1020
CTGCATAACC AA	ACTTGTA	CCAATG	ATAA	TATCCAT	ATC	TCCTTCT	GCC	AACAAG	ACAT	1080
TGATTAATC AT	CTTCATCC	TTTAAAT	TCA	GCGCTT	TGAG	ACCATTT	TGA	CGAATAT	TGG	1140
CAAACCTCCTT AAC	ACTGGTT	CTCTT	CACAA	TACCGT	GACG	GGTTGT	AAAG	AAGAGAT	AAAG	1200
CATCATCACT GCG	ATCAGAC	TCAACAT	TGA	TAACCG	TCTG	AATACT	TTCG	TCTTCAT	CCA	1260
ATTTCAAGAG ATT	GACTACT	GGTAGC	CCTT	TGGCAG	TCCG	ACCATA	CTCA	GGAATTT	TCAT	1320
AACCTTTAAG ACC	GATAGACA	CGTCC	CTTGT	TTGTGA	AGAA	GAGCAG	ATGA	TCATGGG	TGC	1380
TAGTTGACAC TAA	CTCACA	ACAAAG	TCAT	CATCTT	TCAC	TCCCGT	TCCT	TGGACAC	CAC	1440
GACCCCCACG TTT	TTGAGCA	GTGAA	CTCGT	CCTGAT	CCAA	ACGCTT	AATG	TAGCCT	CTGT	1500
TAGAAAGGGT AAT	CAAGACA	TCCGAT	TCTT	CAATCA	AGTC	CTCATC	CCTG	AGACTCA	AGA	1560



1121

CCTGTCCAAT CATCAACTCT GTACGGCGCT TATCAGAAAA TTTACGTTTA ACTTCATCCA	1620
ATTCGTCTTT GATAATTTGA GAAACACGTT CAGGCTTAGC AAGAATATCT GCTAAATCCG	1680
CAATCAGAGC CAAGAGGTCA TCATACTCAG ATTGAATCTT ATCGCGTTCC AAACCTGTCA	1740
AACGACGAAG ACGCATATCA AGGATAGCTT GACTTTGACG TTCAGAAAGC TTAAACTTGC	1800
TCATCAACTC AGCTTGAGCT TCCGCATCGG CTTCACTAGC ACGGATGATA CGAATCAyTC	1860
GTCGATATGG TCTAGCGCAA TCAAGAGACC TTCTAAGATA TGAGCGCGCG CTTCCGCTTT	1920
TTCTTATCA AAACGTGTAC GACGAACAAC CACTTCTTTT TGGTGCTCGA TATAAGCATC	1980
CAAAATCTGA CGAAGAGACA AAATTTTCGG TATACCATTT TGGATAGCGA GCATATTGAA	2040
ACCAAAATTG GTTTGCATTT GGGTCATTTT GAAGAGGTTA TTGAGAATAA CATTGGCTGA	2100
GGCGTCGCGC TTGACTTCAA TAACAAATCG AACACCTTCA CGGTTTGA CTACACGTAC	2160
TGCTGTGATA CCCTCAATGC GTTTTTCCTG AACCAAGCGA ACAATATGCT CATGCACCTT	2220
GGTTTTATTG ACCATGTAAG GAAATCTGT TACAACGATA CGCTCAGCAG CAGTCTTAGT	2280
CGTTTCAATC TCTGTACGAG AACGTAGGAC AATCGAACCT TTACCTGTTT CATAAGCCTT	2340
ATGGATACCT GATTTCCCCA TGACAAGAGC ACCAGTTGGA AAATCTGGTC CAGGCAAGAC	2400
TTCCATCAAG TCCTTGGTAG TCACTTCAGG ATTATCCATG ACCAACTTCA CTGCATCAAT	2460
GTGTTACCCC AGATTATGAG GTGGAATATT GGTGCCCATC CCAACCGCGA TACCAGTTGC	2520
TCCATTAACC AAAAGGTTTG GAAAACGCGC TGGCAAGACC AAGGGTTCCC GTTCATTGGC	2580
ATCATAGTTA TCAACGAAAT CAACTGTATT TTTGTTGATA TCACGAAGCA TTTCCAGAGC	2640
AATCTTGCTC ATACGTGCCT CGGTATAACG TTGAGCGGCA GCACTATCTC CATCCATGGA	2700
ACCAAAATTC CCATGACCAT CTACAAGCAT GTAACGGTAG CTCCACCATT GAGCCATACG	2760
GACCATGGCT TCATAAATAG AGGAATCCCC GTGTGGGTGA TATTTACCCA TGACATCCCC	2820
TGTAATACGA GCAGATTTTT TATGGGGTTT GTCTGGGGTC ACACCCAATT CATTCAATTC	2880
GTAGAGAATG CGACGGTGAA CAGGTTTTAA GCCATCTCGA ACATCAGGAA GAGCTCGCGC	2940
TACGATAACA CTCATGGCGT AGTCGATAAA ACTTGCCTTC ATCTCCTTTG TCAGATTGAC	3000
ATTCATAAA TTTTATCCT GCATTAATAA ATGCCTCATT TCACAATTAG TAAGTAACAA	3060
CATTATACCA TAAATTTCCA TCTATTTTCA CCTCTAAACC ACTAAAACGT TTACATCGAG	3120
AACTATAAGG CATATTCGTG ACAAAGTTTT TTAAGTGA TACAATGAAG TTGTCTAGGG	3180
AAAACCCCTA ATAGAATAAG GAGATGGTTA TACAATGACT CTGACTAACA CACAAA	3236

(2) INFORMATION FOR SEQ ID NO: 181:

1122

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8651 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

AGGTCCTGAA GTATTGGAAC AGGAAGGTCA AGAGTTTTTG GAACATTTCA AAAAACTCTT	60
GGAGTCAGTT GAAGTAGTAG CCATCTCAGG TAGTCTGCCA GCTGGCCTTC CAGTTGATTA	120
CTATGCGAGC TTGGTAGAAC TTGCTAATCA AGCTGGCAAG CATGTAGTCT TGGACTGCTC	180
AGGTGCAGCA CTTCAGGCTG TTCTTGAATC ACCCCATAAA CCAACAGTCA TCAAACCAAA	240
TAATGAAGAA TTGTCTCAGC TTCTTGAAG AGAAGTTTCT GAGGATTTGG ATGAATTAAA	300
AGAAGTACTT CAAGAACCTT TGTTTGCAGG GATTGAATGG ATTATCGTTT CACTTGGTGC	360
CAACGGTACT TTTGCCAAAC ATGGTGACAC TTTCTACAAG GTAGATATTC CTAGAATTCA	420
GGTGGTAAAT CCTGTTGGAT CTGGAGACTC TACTGTGGCA GGAATTTCTT CAGGACTTCT	480
TCACAAAGAA TCGGATGCAG AATTACTCAT CAAGGCAAAT GTCCTTGTA TGCTCAATGC	540
TCAAGAAAAA ATGACTGGTC ATGTCAACAT GGCCAACTAT CAAGCTCTAT ATGATCAATT	600
AATAGTAAAA GAGGTATAAA ATGGCTTTAA CAGAACAAAA ACGTGACGC TTAGAAAAAC	660
TTTCTGATGA AAATGGTATC ATCTCAGCTC TTGCATTGA CCAACGTGGT GCTTTGAAAC	720
GCCTCATGGT TAAACACCAA ACAGAAGAAC CAACTGTGGC CCAAATGGAA GAACTTAAAG	780
TCTTGGTAGC AGATGAATTG ACTAAATATG CTTCATCTAT GCTTCTTGAC CCTGAGTATG	840
GACTTCCAGC AACTAAAGCT CTTGATGAAA AAGCTGGTCT TCTCCTTGCT TATGAAAAAA	900
CAGGTTATGA CACAACAAGC ACAAAACGCT TGCCAGACTG CTTGGATGTT TGGTCTGCAA	960
AACGTATTAA AGAAGAAGGT GCAGATGCAG TTAAATTCTT GCTTTACTAT GATGTAGATA	1020
GCTCAGACGA ACTCAATCAA GAAAAACAAG CCTACATCGA ACGCATCGGT TCTGAGTG/G	1080
TGGCTGAAGA TATCCCATTG TTCCTTGAAA TCCTTGCTTA CGATGAAAAA ATTGCGGATG	1140
CAGGTTCTGT AGAATACGCT AAAGTAAAC CACACAAAGT TATCGGCGCT ATGAAAGTCT	1200
TTTCAGACCC ACGCTTTAAC ATTGATGTTT TGAAAGTTGA AGTTCCTGTT AACATTAAAT	1260
ATGTTGAAGC KTCGCTGAAG GTGAAGTACT TTATACAGT GAAGAAGCAG CAGCCTTCTT	1320
CAAAGCGCAA GATGAAGCAA CGAACTTGCC ATACATCTAC TTGAGTGCTG GTGTATCAGC	1380
TAAACTCTTC CAAGATACTC TTGATTTTGC TCATGAATCA GGTGCGAACT TTAACGGAGT	1440
TCTTTGTGGC CGTGCTACAT GGGCAGGATC AGTTGAAGCT TACATCAAAG ATGGTGAAGC	1500

1123

AGCAGCTCGC GAATGGCCGC ACAACTGGAT TTGAAAACAT TGACGAACTC AACAAAGTTC	1560
TTCAAAGAAC AGCAACTTCA TGGAAAGAAC GCGTGTAAAG AAGTCCTCCT AGTTTAGGAA	1620
CATGAATCTA AAAAAATTTA AAAAAAGTTG TATGTAAAGG CTTACAAAAT AACTTACTTG	1680
TGCTATACTT AAATCACAAG TTAATATGAA TTAGAAAGTA ACTATATGAA GTATAATAAA	1740
AATAGGATAT AGTTTATTTT ACGAGCTAGG AAGGAAAAAT ACGGAAACAA TATTGCCAGA	1800
ATAAACTATA TTTAGATGCA CATTTTCATT ATTGTTTTAT AAAAGGAGAA GATAAACGGC	1860
TACTAAAAAG AGTTTTAAAG CGTTAGTTGT AGGACTAGGT ATTGTTTCAA TATTCCTATC	1920
AGCCTTACCT ATGGTTAGTG GTTCTGTATT TGCAGATAGT GCCCTAACTA CAGTAGATAA	1980
AGCAAATGAT ATTGTTTTGA ATGTTGATGG GAATAAATTT TATAATGTTT CGGTTTCAGA	2040
AGATATTGTA AATGCTGGTC AAATTTTGGA AGATTATTTT TATGTAGATA AATTTGGAAA	2100
TATAAATTTA AAAGGCACTC CTGAAGAGTT AGCAAAAAAT ATTGGTATTT CTGTACAAGA	2160
AGCAAGTTTG ATGTATGGAG CTGTAAAAGA GTTACCCAAC GTTTACGAAA GAGGTCCTGT	2220
AGGTTTTCGT TTCAATCTTG GTCCTCAAGT GAGGGGGATG GGTGGCTGGG CTGCTGCAGC	2280
TTTCGCTACT GGATATGCTG GATGGCATT TGAACAATTT GCGGTTAATC CTGTTACATC	2340
TGGATTTGTT GCTGTAATAA GTGGTGCGAT TGGCTGGGCT GTAAAACTG CTGTAGAAAA	2400
TTATTGGACA GTTGCTGTAG CTACAGTAGA AGTGCCGTTT GTGAACCTTG TTTACACCAT	2460
AGATTTACCT TAGAGGTTAT TTCTTTATGA ATCATTCTTT TAAAAAATA ACTGTATTTT	2520
GTTTTATAGT TTCTGTGTT CTGTGTTTAT TAGACTTAAT GAATTTTAAA AATGTAGCTA	2580
CTTTTTTATT TTTCTGTCTT CCTGTTTTTG TTTTGATTTA CAAAAATAAA TAAAAACAGA	2640
GCCTCTGTTT GATGAATTTT AGAACATAGT TAAGTTTTAA AAAAAAGTTGT ATGTAAAGGT	2700
TTACAAAATA ACTTACTTGT GCTATACTTA AATCACAAGT TAATACAAGG TGAGTGTTAC	2760
TAAGTAATAT TAGGCATGAT CACAGGTGAA TTAGAAATCA GCTGATTTTC TAGTTCATTT	2820
GTGGTCATTT TTTGTACTTA TATACCTTTA AGATATAAAA GGAGGTTGAC ATGTATCGAA	2880
TTCTAAATCC AATGAATCAC AATGTCTCGC TTGTCAGAAA TGATAAGGGA GAAGAGGTGA	2940
TTGTAATTGG TAAGGGAATT GCATTGCGAA AGAAGAAGGG GGATTTGATT GCTGAAAATC	3000
AGGTTGAGAA AATCTTTCGG ATGAAGACCG AAGAGTCCAG AGAAAACTTT ATGGCTCTTC	3060
TCAAAGATGT TCCGCTTGAT TTTATCACAG TGACCTATGA AATCATTGAT AAGCTATCAA	3120
AGAAATATCA TTATCCGATT CAAGAGTATC TCTATGTAAC CTTGACAGAT CATATTTACT	3180
GTTCCTATCA AGCTCTAACT CAAGGAAGGT ACAAGGATAG TAATCTGCCA GATATTTCCG	3240

1124

CTAAGTATCC TGTGCTTTT CAAATCGCAA ATGAAGCTTT TGAAATTAC CGTCAGAAGC	3300
TAGCAGATCA TTTTCCTGAG GACGAAATTA TTCGGATTGC TTATCATTTT ATTAATGCTG	3360
AAGGTGAAAA TGAAGTGGAA CTTGTGGAGT CGATTGATAA GAGGAAAGAA ATTCTCAGGA	3420
ATGTTGAAGA AGTTTAAACG GACTATGCAA TTCAACGAAC TAAAAAGAAT AACCATTCTT	3480
ATGATCGCTT TATGATCCAT TTGAATTATT TCTTGGATTA TTTAGACAGA TCTAGAGATG	3540
ATAACCAATC ACTTCTGGAT ATGGAAGATC ATATTAAACA ATCCTATCCA AAAGCCTTCG	3600
AGATTGGTTC CAAGATCTAT GATGTGATTA CGCAACATAC GGGTCTTGAT TTGTATAAAA	3660
GTGAACGAGT TTATCTAGTT CTACATATCC AACGTTTATT GTCATAAAAA TTTATTTAAA	3720
ACTATATAAG GAGAATTCTA TCATGAATAG AGAAGAAGTA ACATTGTTAG GTTTTGAAAT	3780
CGTAGCCTAT GCTGGCGATG CTCGTTCAA ACTATTGGAA GCCTTGAAGG CTGCTGAAGC	3840
TGGTGATTTT GAAAAAGCGG ACGCTCTGGT AGAGGAAGCT GGTAGCTGTA TTGCAGAGGC	3900
TCACCACGCG CAAACAAGTC TATTGACTAA GGAAGCTTCA GGTGAGGACT TGGCTTATAG	3960
TGTAACCATG ATGCATGGCC AAGACCACTT AATGACAACT ATCTTGTTAA AAGATTGAT	4020
GCATCATTTA ATTGAACCTT ACAAGAGAGG AGTTCAATAA TGAATAAACT AATTGCATTT	4080
ATCGAGAAAG GAAAGCCTTT CTTTGAAAA CTATCTCGTA ATATCTATCT TCGTGCTATT	4140
CGTGATGGTT TCATTGCAGG TATGCCTGTT ATTCTCTTCT CAAGTATCTT TATCTTGATT	4200
GCCTTTGTAC CAAACTCATG GGGCTTTAAA TGGTCTGATG AAGTTGTAGC CTTTCTGATG	4260
AAACCTTATA GCTATTCTAT GGGTATTCTG GCTCTCTTGG TAGCTGGTAC AACAGCTAAG	4320
TCATTGACTG ACTCAGTAAA CCGGAGCATG GAAAAACCA ATCAAATCAA GTATATGTCA	4380
ACATTGTTGG CAGCAATTGT TGGTTTGTG ATGTTGGCAG CTGATCCTAT CGAAAGTGGT	4440
CTAGCTACTG GATTCTTGGG GACAAAAGGT TTGCTTTCAG CCTTCCTTGC TGCCTTTGTT	4500
ACTGTAGCCA TCTATAAGGT TTGTGTTAAG AACACGTCA CTATTCGTAT GCCTGACGAA	4560
GTCCACCAA ATATCTCACA AGTCTTTAAA GATGTGATTC CATTCACTCT ATCTGTTGTT	4620
TCTCTTTATG CTCTTGACTT ATTAGCACGT TATTTTGTG GTTCTAGTGT GGCAGAATCA	4680
ATCGGTAAAT TCTTCGCACC ACTCTTCTCA GCAGCAGACG GATACCTTGG TATTACCATT	4740
ATCTTTGGTG CCTTGCCTT CTTCTGGTTT GTTGGGATTC ATGGTCCATC TATCGTTGAA	4800
CCAGCTATCG CAGCTATTAC CTATGCCAAT GCCGAAGTTA ACTTGAACCT TCTCCAACAA	4860
GGGATGCATG CAGACAAGAT TCTTACTTCT GGTACACAAA TGTATTATCGT TACCATGGGT	4920
GGTACAGGTG CGACATTGGT CGTTCCATTT ATGTTTATGT GGTGACAAA ATCGAAACGT	4980
AACCGTGCAA TCGGACGTGC TTCAGTAGTT CCTACCTTCT TCGGTGTAAA TGAACCAATC	5040

1125

TTGTTTGGTG CACCTCTTGT TTTGAATCCA ATCTTCTTCA TTCCATTTAT CTTTGCTCCA	5100
ATTGCAAACG TATGGATTTT CAAATTCTTT ATTGAACTC TTGGAATGAA CTCATTCACT	5160
GCTAATCTAC CATGGACAAC TCCAGCTCCA CTAGGTCTAG TTCTTGAAC TAACTTCCAA	5220
GTGCTATCAT TCATTCTTGC TGCCCTTCTA ATCGTGGTTG ACGTTGTCAT TTACTATCCA	5280
TTCTTAAGG TCTATGATGA ACAAATTCTT GAAGAAGAAC GTTCAGGTAA GTCTAATGAT	5340
GAATTGAAAG AAAAAGTTGC TGCAAACCTC AACACTGCAA AAGCGGATGC TATTCTTGAA	5400
AAAGCGGGTG TCGATGCAGC ACAAATACC ATCACTGAAG AAACAAATGT CCTCGTTCTC	5460
TGTGCAGGTG GAGGAACAAG TGGTCTCCTT GCAATGCTT TGAATAAGGC AGCAGCAGAA	5520
TACAATGTCC CTGTGAAAGC AGCAGCAGGC GGCTATGGTG CTCACCGTGA AATGTTACCA	5580
GAGTTTGATC TTGTTATCCT TGCCCTCAA GTTGCTTCAA ACTTTGAAGA TATGAAAGCA	5640
GAAACAGATA AGCTCGGTAT TAACTAGCG AAAACAGAAG GCGCTCAATA CATCAAATTA	5700
ACTCGTGATG GAAAAGGTGC TCTTGCAATC GTACAAGCGC AATTCGATTA AGGCTAGAGA	5760
CTCTGAAATA GTCTCCCATC GTTACGGAAA TCGCTATGGC GAATTCCTTA TTATTAATTC	5820
GTCGGTAAAA AGATATCGTT TTTACCTCCT CATGTCACAA TTCGGTGAAT TGGTACAAGA	5880
AGTGAGATGG AGAAGGATGG CTCACTGACT CCTCTCCTCT CACTTTTACT TTATTTAAAT	5940
CAAGAAATAG GTGAAAAAAA TGACAAAAAC ACTTCCAAAA GACTTTATTT TTGGTGGCGC	6000
AACAGCTGCT TATCAAGCAG AAGGTGCTAC ACATACTGAT GGAAAAGGAC CAGTTGCTTG	6060
GGATAAATAT CTTGAGGATA ACTACTGGTA CACTGCCGAA CCAGCTAGTG ATTTTACAA	6120
TCGATATCCA GTTGACCTCA AGCTAGCAGA AGAGTATGGT GTCAATGGTA TTCCAATTC	6180
TATTGCTTGG TCACGTATTT TCCCGACTGG TTACGGCCAA GTAAATGCTA AAGGTGTTGA	6240
GTTTTATCAT AATTTATTTG CAGAGTGTC AAAACGTCAT GTTGAGCCTT TTGTAACCTC	6300
TCATCACTTT GACACGCCAG AAGCTCTCCA CTCAAATGGA GACTTCTTAA ACCGTGAAAA	6360
TATCGAACAT TTTGTAGACT ACGCTGCCTT CTGTTTTGAA GAATTTCCAG AAGTAAACTA	6420
TTGGACAACC TTTAATGAAA TTGGACCAAT CGGTGATGGT CAATATTTGG TTGGGAAATT	6480
CCCTCCAGGT ATCCAGTACG ACCTTGCCAA AGTCTTTCAA TCACACCACA ATATGATGGT	6540
GTCTCATGCA CGCGCGGTAA AATTGTACAA AGAGAAAGGC TATAAAGGGG AAATTGGTGT	6600
TGTTACGCCC CTGCCAACTA AATATCCTCT AGATCCTGAA AATCCAGCAG ATGTTCTGTC	6660
AGCTGAGTTG GAAGATATCA TCCACAATAA ATTCACTTTA GACGCAACTT ATCTAGGTGC	6720
CTATTCAGCT GAAACCATGG AAGGTGTCAA CCATATCTTA TTAGTCAATG GTGGTAGTTT	6780

1126

GGATCTTCGT	GAAGAAGATT	TTACAGCATT	AGAAGCTGCA	AAAGACTTGA	ATGATTTTCCT	6840
AGGAATCAAC	TACTATATGA	GTGACTGGAT	GGAAGCCTTT	GATGGAGAAA	CTGAAATTAT	6900
CCATAATGGT	AAAGGTGAAA	AAGGAAGCTC	TAAGTATCAA	ATCAAAGGTG	TTGGTCGTCG	6960
TGTAGCTCCT	GACTATGTAC	CACGCACGGA	TTGGGATTGG	ATTATCTACC	CTCAAGGTTT	7020
GTATGACCAA	ATCATGCCGTG	TGAAGAAAGA	TTATCCTAAC	TACAAGAAGA	TTTACATCAC	7080
TGAAAATGGT	CTCGGCTATA	AAGATGAGTT	CGTTGATAAC	ACTGTTTACG	ATGATGGTCG	7140
TATTGATTAC	GTGAAGCAAC	ACTTGGAGGT	TTTATCTGAT	GCGATTGCAG	ATGGAGCTAA	7200
TGTAAAAGGT	TACTTCATTT	GGTCATTAAT	GGATGTCTTC	TCATGGTCAA	ACGGTTATGA	7260
GAAACGTTAT	GGTCTCTTCT	ACGTAGATTT	TGAAACTCAA	GAACGTTATC	CTAAGAAATC	7320
AGCTCACTGG	TACAAGAAAG	TAGCGGAAAC	TCAGATTATA	GACTAGTAGA	ATTAGTCATT	7380
AGATATAGAA	TTTTAGTGAG	TCAAAAAGAT	GTTCAAAGAT	TTTATCCAAT	CTATTTATGA	7440
AAAAAAGTTT	ATATTATAAA	TTTCGAAAAA	TGCTCTCAAA	TACCGTGTTT	GACGAGTGAA	7500
GAATTGAAAA	GTCTTGAAAA	ATGGTATGTC	TCGACTGGTA	AAGAATGGAT	TTGTCATTCA	7560
GATGATGAGC	TGGAAGAATT	TAAAAATCTA	TTTTTAAATT	TTATCAATCC	TGAAGAATGG	7620
GATACTATCT	CCTTTGATTC	AGATTTTATG	CCGTTTCAAC	AATCGTAACC	AATTTCTCAA	7680
AAAAGTTAAA	TCTTATATTT	AGTACTCTGT	AAAACCTCTA	TCTAATCAGC	TTGCTTATAC	7740
TCAATGAAAA	TCAAAGAGCA	ACTTTAAACT	AGGAAGCGAG	TCGCAGATTT	CTCAATGCAT	7800
AGCTTTGAGG	AATTGGGCAA	AAAGTCTTTG	ATATAGAAAA	ACGCATAGTA	TCAGGTGTTT	7860
CAACACCTGA	TACTATGCGT	TTTATGTGG	GAAGATTTAC	TTTTTTTCTT	CTGAAATTGA	7920
GTTGTTACCC	AGGCTCTTTC	AGTTTATTAA	GGCTTGATGA	CTTTAATGTG	TTTAGATAGC	7980
TTAAAAAGGA	TTGAATCACT	TAGTTTAGAA	TCTGAAACAA	TAGTATCAAG	ATTTGATACA	8040
TTATAAAAAG	TATAAAAATC	AAACTTATTG	AACTTGCTAT	GATCTGCGAG	TAAATATTTT	8100
TTATTAGAAT	TATTTAAAGC	GATGCGTTGA	GCCTCTCCCT	CTTCTCGCT	AAAAGTAGLT	8160
AGAGCTCCGT	TTGAATACC	ATTACAGCTA	ACGAAAGCTT	TAGAAAAATTG	GAGATTAGAG	8220
AGATTTTGTA	GGGTCAATGT	ACCAACAAAA	GCACCTGTAA	TATCGCGATA	ATTTCCACCT	8280
ATTAAAAATCA	AATCTGTAA	TTTTCGTTCC	CTTAAAAATCA	GAAAAACAGG	TAGACTGTTG	8340
GTTACGACGC	GGATATTGTC	AATAGGCAAC	TCACGCGCAA	AAAACCTCTAA	TGTTGTTTCT	8400
GGTCCAATGA	AAATAGTTTC	TCTTTCTTCT	ACTAGACTGC	CTGCAAAATG	GGCTATTTCT	8460
TGTTTTTCTG	CCGTTTGGAG	GGCTTGTTTT	TCAATATTTG	ATCGCTCATT	AGTCAAAAGG	8520
GAGTTGGTTC	GAAGTTTTTC	AGCTCCACCA	TGCACACGAA	TCAGCAAATC	TTTATCAGCT	8580

1127

AATTCCTGTA AATAGCCCT TGCAGTCATA TCTGAAACGG CTATTTCTGC CATAATCTGT 8640  
TTAACTGTTA T 8651

(2) INFORMATION FOR SEQ ID NO: 182:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1786 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

AATCTCCAAT CAGTGCCACT TCAGCTACAA AGAAGAGGAG GATAATAACT CCGTTCACAA 60  
GGACAGACAA GAATAATTGA TAGAAGGAGT CGGTTTCACT TGCTTGACTT GGTCTTGTA 120  
TGATWTGGAG ACTGGCAAGC AGAATGATTC CAATGCTAAT CACACACAAG AGGGCTGTAA 180  
ATCGTAGGCT ATCAAAGAAA GCAAAGAAAC TAGCAATAGC AGTGAGGAGG ATTGGAATTG 240  
CCAAGAGTTG ACTATATTGT TGGAGAACCT TGTCTAGCGT CCAGTCCTTT TCCTGGTGGA 300  
TAAATCGTCT CACAACGAAA CTACCCAAGA GGAATGAAAA GAAGAAGACT GTTGTGCTA 360  
CTAGGATAGA GATGATAGAA AAAAGAGTTA AAGGAGCTAG CTGCTCAGGG AAGCGACTGT 420  
TAATGCTTGC TATATGTCCA TAGTAAGCAT GTTTGATGTG ATAGATACTA AAGAAAAAGG 480  
AAGATGCAGA AAACAGAAATG AGCAAGAGAA AGGCTGTGTA ACTGTGTGTG AACTTGTGTT 540  
CCAACCTACT TGTAGGAGAT TTGATCGCTT CCACTAGCCA AGACCAAAAA TCAAGCACTT 600  
GCTCTTTCCA TTTATCCCTA GATTTTGGAG CTTGGTCCGG GATATAAGGA CTTTCTAAAG 660  
ATTTACTGAT AAGAAGTGGC TCTTTCGTGG TTGCTTTTTG CTGAGGAAGA GCTTCTTGGC 720  
TCTCTTCAGC TATAGTGAAT TTTTCTGTTT CTTTAGAAAG GTCTGGCTCT TCTTCAGTAG 780  
AATTAGATGC CTTCTTTTCT TCTATTTCTG TTCTCGCTTC ACTGTCTTCA GGAGCTTCAA 840  
TTTTCTCTTC TTGCTGGCTT TCCAATTCGA CTTCACTTG AGGGACTTCC TCCTCTAACT 900  
GAGTATTTTT TTCAATTGGT GTATCGAGAT CGGCTATCGT TTCTTCAGCC TTGCTGCAA 960  
CCTCTTGAGC TTGCTCTTCA GGCTTGTCTT TGCTTGTGT TTTTACAAAA TCATTACTTT 1020  
CAAACCATTC TTGTTTCATG GTAGAACCTC CTTTTTAGTT AGATAAATAT GTTTCATAG 1080  
TAGCAAATGT AAGCGTTTTT GTCAACGTCT GCTTGGTGTG GATATTAGAT CAATATTATC 1140  
ATCAGATCTC GCAATGAGTT GATCCTTGAC ATCGGTTTTT TCAGTTTTGT AAGGGTTGCT 1200  
TAATCCGTA CCTCTTGATT CAGGCTTTTC TCTTGTGAAT TGGAAGATAG AACCATAGTT 1260

1128

GCTTGAGATG TCCCAGTTAA TTCGTTGGCT TTCTTTCTGG TCTAGGATGA TTCTGAGATA	1320
ATCTTTGGCA GTCAGTTCAA CCTTGCCATG GACTTGGATA TTTTCAGCGT GGAAGTGATT	1380
CTCTGTTGAC TCTAGCTGAC TATCTGTAAG AACTGTATCA AAGATATTAA CGATATTGGG	1440
CGTTGTGAGT TTA CTGTTTT TGATACGACT TCCTTCAATT CGGAGGATAT AGCTGTTTGT	1500
ATTGAGGGTC GCATTTTCAA GGCTAGCATT TATGATGGTG GTTTGTCCGC GATTGGCTGA	1560
GATGTTGATC CCTTTTAGAG TTCTCCCTTT TGGTAGTCGG AGAATAACTT CTTCAAAACG	1620
ACTAGAGTAG CTACTTGCGA TATGAAGAAT CCCACCAATT CCAGAAGAGA GAAACGGAGT	1680
TTGAGACAGT TTCTTATCAG TGAGACTCAG AGTTCTATCG TTCTGATTGG TGATAAGATC	1740
ATGGTGAGCA GAAAGAGATG GATGGTAAGA AATGTGGATT TGATCATCGA AAGAGTCTGT	1800
GATGGTGAGC GTGTGTTGGT GGAGAGTAAT TTCTAGGTTT TCGACTTCCT TGCCAAAGGT	1860
TAGCTTTTCC GTACGGCTAT CATAGACAGG TTCTTTGGAC ATGGAAAGTA GGCTCTTAA	1920
CCCGTCAGAT TGGATACCTA CAAAAAGCAG GATAAAGCCG ATAACGGTAG TCACCACACC	1980
AAAGATGAGA AATCCTTTTG TCCATTTACG CATGCTGATT ACCTCTCTTT CCTTTTAA	2040
GAACAAATTG TACCAGACGA ACAATGAGTA GACCGAAGAA GCGAGTTGCA TAGGAAATGC	2100
CAAGTAAAC TAGCGAAGAA GCACCGATAG CCAGTAAACC AGAACCAAAA ATCAAGATAA	2160
AGGCTGATTT GGCTTGGCG AGGACAGTGA AACTTTCAAC TAAAAATAGG AATCCGCCGA	2220
TGATACCCAG TATGGAACT GCAAAGAAAG CCAGAATGAC AGTCAAAGCG GCTACAAGAA	2280
TTGCGAACAG GGTCACGAGG ATGGCGATTC CCAGAGGAAT GCCGATAGGT GCTGCAAGGA	2340
GGGCTAACAA GCGATATGT AAAATTTGTC GGTATTTTT TTGAGCGGGT GCTTCATTGA	2400
TTTTTTATC GAGAAGATTG GATAGAACTT CGTGGGCCGC TTCTTTGGGA GTTCCCAAAC	2460
TACCGATGAG TTCTTCTTCT CCTTCGACTC CAGCATCGTC AAAGAGCTCT CTGAAATAGT	2520
CCATGGCTTC GATACGGTCA GCTTCAGGTA GTTCTTGAG ATAGAGTTCT AGCTGAGTCA	2580
GGTATTGAGT TCTGTGATG GCGGATACTC CCTTCTATGA TGCCATTGAT GGTGTCTGTA	2640
TAGAGTGCCC ATTCATCTTT TAGGGTCAAG AGCTGCTCTA TACCACCGTT TGTCAAGGAG	2700
TAGTATTTGC GCATGCGACC TTGGAATCT CTAGAATAGG TTGTCAGAAA GCTATTGCCT	2760
TCCAATTTTT TGAGAATGGG ATAGAGTGTG GATTCTTTGA TATTAGCGAT CAGCTTAATG	2820
GTTTGGCTAA TCTCATAACC ATAAGAATCA CCCTGCTCCA GTACAGCCAA GATGAGAAAT	2880
TCAATCAAGG CAGAGGATGT TGGAAAGTAC ATGGGAAACC TCCTTTTCTA ATGTGTAAGA	2940
TTTTTATATA TAATTTTTCT ACACATACAT TGTACATCTA AAAGAAAGCC CTGTCAAGAG	3000
AAATGTGTAA AATTTTTATA TATAAAAAAC TTCTAGCTAA AACTAGAAGT TTAAAGGATC	3060



1129

TTATCCGCTC TGTCCACTGT AAAGAGGGCC ACAGTCATCA GGATATCGAT GAGCAAGAGG	3120
GCAGCTACAG ATGGTACCCA AGAGTGGAAC AGGTCAAAAC TGTAACCAA GAGGGTTGGC	3180
CCAAAGGCTG CTAGGATATA GCCTCCTGTT TGAGATAGGC CGGACAATTG GGCTGTCTTT	3240
TCAGGGGCGC TTGTCTTGAG TGAAAAGTTG ACCATGAGAT AAGGGAAGAG GGCACGTGTT	3300
GCGGTTCCGA TGAGGAGATG GATGGCAAGC CAGTAAATGA AATTATTGAT TGGGAAAAAG	3360
AGCATGGAAA TGCCGACCAC ACCAGCTAGT GAAACCAGAG TGAGCATGAG CTGACGGTTG	3420
CGAGTAGATA AACTGGTTGT CAGGCTTGGG ATGGTCATTG AAAAAGGAAT GCTAATCAGA	3480
GATAAGATAG AAGTCAGCAA GCCAGCTTCG TGA CTGGATA GACCTGCATG GATAGACATG	3540
GTAGGTAACC AGGTCATGAC GGTGTAAAAG ATCAAGGATT GAAAACCTGA AAAGATAATA	3600
ATTGCCCAAA CCTGTTTATT ACGCATGACC TTTATTTGAC TTTTTTGT TTGTTTGTGA	3660
GCTAGTCTAT GATTATAGCG GTGATTTGGG AGCCAGACCA AAAAAGTTGC TAGACAGAGT	3720
AACGTGAGGA GAAGGATAAG TCCTTTCCAA GAACTGGCTT GTGTAATGGG CACAGCTAGA	3780
TAGGAA	3786

## (2) INFORMATION FOR SEQ ID NO: 183:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3054 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

TCAGCTAAAA AACATTGCTA AATTGATTGA AGCTGGTGCT ACACATTCCG ATTCAACTTC	60
TCACACGGCG ACCACCAAGA ACAAGGTGAG CGTATGGCAA CTGTTAAACT TGCGGAAAAA	120
ATTGCAGGTA AAAAAGTTGG TTTCTTCTT GATACAAAAG GACCTGAAAT CCGTACAGAA	180
TTGTTCGAAG GTGAAGCTAA AGAATATTCA TACAAAAC TGAAAAAAT TCGTGTGCA	240
ACTAAACAAG GAATCAAATC AACTCGTGAA GTGATTGCGT TGAACGTTGC TGGTGCTCTT	300
GATATCTATG ATGATGTTGA AGTTGGTCGT CAAGTTTGGG TTGACGATGG TAAACTTGGT	360
CTTCGTGTGG TTGCTAAAGA TGATGCAACT CGTGAATTG AAGTTGAAGT TGAAAAACGAT	420
GGTATCATCG CTAACAAAAA AGGTGTGAAC ATCCCTAACA CTAAATTC TTTCCAGCT	480
CTTGCTGAAC GCGATAACGA CGATATCCGT TTCGGTCTTG AACAAGGTAT CAACTTCATC	540
GCAATTTTCAT TCGTACGTAC TGCAAAAGAT GTGAACGAAG TTCGTGCAAT CTGTGAAGAA	600

1130

ACTGGAACG GACATGTTCA ATTGTTGCT AAAATCGAAA ACCAACAAGG TATCGATAAC	660
TTAGATGAAA TCATCGAAGC AGCTGATGGT ATTATGATTG CTCGTGGTGA TATGGGTATC	720
GAAGTACCGT TCGAAATGGT TCCAGTTTAT CAAAAATGA TTATCAAGAA AGTCAATGCT	780
GCAGGTAAAG TTGTTATCAC TGCAACAAAC ATGCTTGAAA CAATGACTGA AAAACCACGT	840
GCAACTCGTT CAGAAGTATC AGATGTATT CACGCTGTTA TCGACGGAAC TGACGCTACA	900
ATGTTGTCAG GCGAGTCTGC AAACGGTAAA TACCCACTCG AGTCAGTAAC TACAATGGCT	960
ACAATCGACA AGAACGCTCA AGCTCTTCTT AATGAATACG GACGCTTTGA TTCAGATTCA	1020
TTTGAGCGTA ACTCTAAGAC AGAAGTAATG GCTTCTGCTG TTAAAGATGC TACTAGCTCA	1080
ATGGATATCA AATTGGTTGT AACTCTTACT AAGACAGGTC ATACTGCACG TTTGATTTCT	1140
AAATACCGTC CAAATGCTGA CATCTTAGCA TTGACATTG ACGAATTGAC AGAACGTGGC	1200
TTGATGTTGA ACTGGGGTGT TATCCCAATG TTGACAGATG CTCCATCTTC AACTGACGAT	1260
ATGTTGAAA TCGCTGAACG TAAAGCGGTA GAAGCAGGTC TCGTTGAGTC AGGCGATGAT	1320
ATCGTTATCG TTGCTGGTGT GCCAGTAGGA GAAGCTGTT GCACAAACAC AATGCGTATC	1380
CGCACAGTAC GTTAAGAAAA ATATAAAAAC CTATCATATC CAGCTTTAGA GCTTGTGTGA	1440
TAGGCTTTTT GTATAGAGGG TAAGAAATAG GCAAACTTT CATAATGGAT TGATACTCTT	1500
CGAAAATCTC TTCAAACCAC GTCAGCGTCG CCTTACCGTA TATATGTTAC TgACTTCGTC	1560
AGTCTATCT ACAACCTCAA AGCAGTGCTT TGAGCAACtG CGGCTAGCTT CCTAGTTTGC	1620
TCTTTGATTT TCATTGAGTA TGAAATAAGA TATGCACAAA TTGATTAGAA AGTCAAATGA	1680
ATTTCTACAA ATGTTTtagc AATCGTAATG TACTTGTCTA GATTCGATCT GATATATTTT	1740
CGATTTAATG ATATGGTATT TAAACCTCC AAAGTAGCTT ACTCCATTCT TTTACTTACG	1800
TGAGTGTAGA TGTTATTTAC TGTTTTAGCG TTTTGTGTT CCACTCTAAC CATTATAGCA	1860
TTCTTCTCAG CTAGTGTACT AAGGAGTGTG TGCCTGAAAA TATGGGAACT AAGGGGCTGG	1920
TTTATCGGTT TCTCTAGTTT AGTATTTGCC TTTTGCAAAG TGATCTTAAA TGCCTTTCTC	1980
TAAATTTACA TATCACTATT GTTTAACAAA ATCTAATCTA TTTTAGGTCA CTTATTCTTT	2040
TTTTGAAATG TAGAATGAAC TTTTCAAAG TTTTCTGAAT CTTTTAAAT CTGTTTGCTT	2100
TATATCGCCA TTCTCCCCC TTTTTAATT CTCCTATAT AGCCTGACAG CTTCCCGAT	2160
GGTACGAATA TGGTTGCTTT CGTCTAGGTG GATGTCGGGG TATTCGGGAT TGAGTTTTTT	2220
TGAGGCAGCC TTGGCGGAGT TTCTTGACAT AGTTAGTGCC GTCTACTTGG AAGATGCCGA	2280
TGGTATTATA GTCAATCTGT GGGGTATTCT TGATAAATAG GTAGTCGCTG TTTCTTATCT	2340
TTGGCTCCAT GGACTTGCTG ACGACATAAG CGATTGGGTC GTAGTCGTCT GGGATAATGG	2400

1131

AAACTCCATA TCTAAATCGT TGTCTGTCAT CGAGCGGCTA CCTGCAGAGA TAAACTACCT	2460
AACACGAGAG TAAGTAGTCT GTCTGTAGTC GTCCAGTCTG ATGATTTTTA CGATACTTCG	2520
TTTTTCTGAT CATAcAGTTG CCTCTCGGCA TAGGTCAGAA CTTTACCTTG TCTGGGTGGT	2580
TCCCGTTGGT CGTAGATAGA TTGGATATCG CTAGGAGAAT CCTTTTGAAC TGGAGGAAAG	2640
AGGGCATCGA TCAAGCTACT GAATACTTTA ACTAAGTCAA ATATAGTATT TTTCTTAGTA	2700
GACCTAACCC TTTTTCATA ATTTCTAATG GTGTTTTTAC TTATACCTAT CTTAGTACCC	2760
AATCTTATT GAGTCCAACC ATTACTAGTC TATATTGTTT TATAGTTGAT TGAGTTTGA	2820
ATAGTACGCT GTAGCTGCTA AAACATTCT AGAAATTAAT TTGACTTTCC TAATAGAGTT	2880
GTTCATATCT TATTTCAATC TATTATGTTT TTCACCTCTA ACAATCGCAA TCTCTCTTT	2940
ATCCATGAAT GAAATCGCTT TCTATTTTGT TAAGTAAAGC ATAACACGAA ATCCACGAAA	3000
ATGAAAACCT TTGTTGTGTT TTCGTAAAAA ATTTGTTGAC AGAGCACGAA ACGC	3054

(2) INFORMATION FOR SEQ ID NO: 184:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1590 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

TGTGATTTTC yGAAAATTTG GTAAATATA TCTTAATCAT TTTCAGGAGG ACAAAAATTT	60
GACAAGATAT CAGAATTTAG TAAATGGAAA ATGGAAATCA TCTGAACAAG AAATTACCAT	120
TTATTCACCA ATCAATCAAG AAGAATTGGG TACAGTTCCA GCCATGACTC AGACTGAAGC	180
TGATGAGGCT ATGCAAGCTG CGCGTGcAGC CCTGCCAGCA TGGCGAGCTT TATCAGCAGT	240
TGAACGTGCG GCTTATTTGC ATAAAACAGC AGCTATTTTA GAACGCGATA AGGAAGAAAT	300
TGGTACTATC CTTGCCAAAG AAGTAGCAAA AGGGATTAAA GCAGCAATTG GAGAAGTAGT	360
GCGTACAGCA GACTTGATTG GTTATGCTGC TGAGGAAGGT CTCCGTATCA CTGGACAAGC	420
AATGGAAGGT GGTGGTTTGG AGGCAACAAG TAAAAACAAA CTGGCTGTTG TCCGTCGTGA	480
ACCAGTTGGT ATCGTGCTAG CGATTGCTCC CTTTAATTAT CCAGTTAATT TATCTGCTTC	540
TAAAATTGCA CCTGCCTTGA TTGCAGGAA TGTGGTCATG TTAAAGCCAC CAACACAAGG	600
TTCCATTCTT GGA CTCTTGT TGGCTAAAGC ATTTGAAGAA GCAGGGATTG CGGCAGGTGT	660
TTTCAACACC ATTACAGGTC GTGGTTCAGA AATTGGGGAT TATATCATTG AGCACAAAGA	720

1132

AGTCAACTTC ATCAACTTTA CAGGTTCAAC TCCTATTGGA GAACGTATTG GTCGTTTAGC	780
TGGTATGCGT CCTATCATGT TGGAACTTGG TGGGAAAGAT GCAGCTCTTG TACTAGAAGA	840
TGCAGATTTG GAACATGCTG CCAAGCAAAT TGTTGCGGGA GCCTTTAGCT ACTCAGGACA	900
ACGTTGCACG GCCATTAAAC GTGTCATTGT TCTCGAAAGT GTAGCAGATA AATTAGCTAC	960
TTTGCTTCAG GAAGAAGTTT CTAAATTAAC AGTTGGTGAT CCATTTGACA ATGCTGATAT	1020
TACACCTGTT ATTGACAATG CTTCAGCCGA CTTCAATTTG GGCTTGATTG AGGATGCACA	1080
AGAAAAAGAA GCTCAGGCTC TTACACCAAT CAAACGTGAG GGCAATCTTC TCTGGCCAGT	1140
GCTTTTGGAC CAAGTTACAA AAGATATGAA AGTGGCATGG GAAGAGCCAT TTGGTCCTGT	1200
TTTACCAATC ATTCGTGTGG CTAGTGTAGA GGAAGCTATT GCCTTTGCCA ACGAATCTGA	1260
ATTCGGCCTT CAATCATCAG TCTTTACAAA TGATTTCAAA AAAGCCTTTG AAATTGCTGA	1320
AAAACTTGAA GTAGGTACAG TCCACATTAA TAATAAAACC CAGCGTGGTC CAGATAATTT	1380
CCCATTCCCT GGTGTCAAAG GTTCTGGAGC TGGAGTGCAA GGAATTAAAT ATAGCATTGA	1440
AGCGATGACA AATGTCAAAT CCATTGTTTT TGATGTGAAA TAACGTGTAA AACCAGGAAA	1500
TTGTTTTCCCT GGTTTTATTT TTTTGCTATA AAATAATAAT AATTATAGAA AAAATACGAA	1560
CTTTTTGGTA TTATAATAGA TTGAAACCGG	1590

## (2) INFORMATION FOR SEQ ID NO: 185:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4848 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

CCTGCAGTTG TCAGACCTGT AATTTTCTTT TTATCTGTAA TAAGAATCGT TCCAGCGCCT	60
AGAAAACCCA CACCTGATAT AACTTGAGCT CCTAATCGTG TAGGATCTCC TGTCCCAAAT	120
TTATAAGATA CGTATTCATT CGTCATCATA ATCAAACATG CAGCTAGACA AACAACTACTA	180
TAAGTTCGGA TGCCTGCAGG CTGGGATTTG CTCCTCTCT CTAAACCAAT TATACTACCA	240
ATGACTACTG ATAAAACAAT CCTGACAACT ATTTCAATAT TTGATAACCC AAGACTAGTG	300
GCTGTCATGA TTATTTCCTT ACTTTACGCC CCGGTCTTTG TGTGAAGTAT AATACCGTTC	360
CAGAAATAAT CATCAGAACA ATTGATATAA CAAATACCAG AGCTTGTGCA TTAGATGTTG	420
CTGTTTCATC ACCTGCAGAT CGAATCGTAA TACCTAATGG TTGAGCTAGG GGATGGTAAA	480
GGAATACAGA TAAGTCGAAG TCAGTTAATA AAGAGTTAAA GTTTAAAGCA ATAACAGAGA	540

1133

GAACAACCGG TAAATATAAT GGAATGATAA CCTTCATCAT AGTATAAAAA GGTGAAGCAC	600
CCATACTTCT TGCTGCATCT TCCATCTCAT CATCAACACT AAATAAAATA GCACGTACCA	660
TTCTATAAGA AAATGGGATT TTTACAATA TATATGCAAT AAGTAGAATT ACCAAACTAC	720
CTACCAAAAT CTGATTCAAG ACAAGAAATT GTGGCTGATT AAAAGTAAAT AATAAACTTA	780
CTGCTAAAAG TGTACTTGGT AGTAACCAAG GAAGTAGAGC ACCATATTCA AATAAGAAAT	840
CAAAACGAGA TTTATGTTTT CTGACAACAC GAGCAAATAC AACTGCGAGA ATTGTTGCTG	900
TTGTGCGAGC AATAATAGAA TAAATAAAGC TGACCAAGAA TGGAGAGAAT GCCGCACTAT	960
TACTAAAGAA TAAGCGATAA TTTTCTAAAG TAAAGTTTGA TAATGTTAAG TTACCTGTTT	1020
GAATTGCAAC TGGATCTGTA AATGAGTATA ATACTATAAA AATTAGTGGA AGCATGAAAA	1080
CTGTGAACAA TCCATATGCT ACAATGTGAG CAATGATATT CCAAGGCTTA GACGCAATTT	1140
TTTGTTTTTT AAGAGGCGCT TTAGTCTTAG AGATAGAAAT ATAATTCCA CCTTTTCTA	1200
TCTTATTCAT GATAGTAAGC AAAATTGTAG TTGCAATACC TAAAATAATT GCAAGTAGGG	1260
CAGCTAAATC ACGAGAATTC CCCATCCCTG CAAATGTAAT AATCATTGGA TTTATAGTTT	1320
GAAATCTTTT ACCACCAACA ATCATGGGTG CTGCTACTGC AGATAAACCA CTAAGAAAAA	1380
CCATAATAGT AAGTGCAAAT AGAGTTGGAA TTAAGGTTGG TAACACTACT TTTCGGAAAA	1440
CAGTAAATGG TTTTGCTCCC ATATTTGAG CAGCCTCAAT AGTGTGATAG TCAACGCTTC	1500
GAATTGTATT TGTTAAAAAC AATGTATGAT TAGCAGTTCC TGAAAATGTC ATAATGAATA	1560
AGACTGCACC ATACCCAATA AACCAGTTAG GGTCTAAAGA AGGGATAACA TTTTGTAAAA	1620
ATTTTGTAAT CAATCCATAA GGACCATAGA CAAATTTATA TCCAGTCGCT AAAACCACTC	1680
CTCCATAAAT TAAAGAGGTC ATATAACCTA ATTTTAAAAT TTTAGCACCT TTAATATCAA	1740
AGTACTCTGT AAATAGAACA CAAAGAATAC CTACGACATT AACTGTAATA ATGAGTGAAA	1800
ATGCTAACTT AAAACTGTTC ATAATACTCT GAAGTGCCCT CTGAGATTTT AGAACACGAT	1860
GTACAGCATC AAGGGAAAAT TCTCCTCCTT TTACAAATAC ATTCACTACT AGATCAAAGT	1920
TTGGATAAAT AATAAATGTT ACTAAGAACC AGATTAACCC TAAACGAATA AGCCAATCTT	1980
TTAAATTTAA TTTATGACGC ATACTGCACC TCCTTAAAAT TGCAGAACGT CTGATGGTGT	2040
GATAAATAAT TCCACACTTT CTCCGACAGA TCTAATAGCA GCCTGACTAT CAATACTTGT	2100
TACATTAAGA ATCTGACTTT CAGAACTTT TATTGTATAG TGAATTGTAA CTCCAGAAAA	2160
CTCAACATCA ATAATTGTCC CTTTTAGAAT AAAATCTTGT TCAGTTTCAC GATTGAATCG	2220
AACTTTCTCT AATCGAATGT ATCCTTTTTT ATCCTCTAAG AAAACGCTTG TATTTTTCAA	2280

1134

TAATACTTCG TGGACTGTTT CATCGGTCAA AACATTAATA TCTCCAATAA AATCACATAC	2340
AAATTCAGTT TGAGAATTAT GATAAATCTC TACTGGTGTA CCGACCTGTT CGATGTATCC	2400
ATTGTTAAAG ACTGCAATTC TATCAGATAA AGTCAAGGCT TCCTCTTGAT CATGAGTAAC	2460
ATATAAAGTA GTAATACCTA ACTCTTTTGT AAGTCTTTTC AACTCTTTTC TCAAATCTAC	2520
ACGTAATTTT GCGTCAAGGT TTGACAATGG TTCATCTAGA CAAAGAATTT TAGGTTCAAG	2580
AACCAGAGCA CGAGCCAATG CTACCCTTTG TTGTTGACCC CCAGATAATT CTGATACATT	2640
ACGCTGTAAC TGTTGATCAG AGATCTTAAT TTTTGCTGCC ACTGCTGATA CTTTAGCTTT	2700
AATAACATCT GGAGCTACCT TCTTAACTTT TAAACCAAAT GCAATATTAT CAAAAACAGT	2760
CATAGTTGGA AATAGCGCAT AAGATTGAAA TACAATACCA ATTCCACGCT TTTCAGGTTT	2820
CAATGAGTG ACATCTGTTT CATTAACTTC AATACTTCCT GATGATGGAT CTAGAAAACC	2880
TACCAATGCT CTCAAAGTAG TTGATTTACC ACATCCTGAA GGCCCAAGAA ATGTAAAAAA	2940
TTCCCCTTCA TGTATATCTA AATTCAGATT ATCAATTGCA ACAAATCAC CATATTTAAT	3000
TTGAATATTA TCAAATTTAA TCATCTCACT AACTCCCTCT ATTACTAAAC CAAAAGCCTC	3060
TCTTTATTTT TTCCATAAAT TTAGAAATAA TAGAGAGACT TGGACATAAA AATTAACCTT	3120
TATTTCTTAT TGTACGTATT CTAATTCAGC TTTTCTACC CATTCAATCA AATGCTTTCC	3180
AACAGCTTCC CAGTCAATAT TTTGTGGTTT CACTTGATCA ACAAATTTCT TCGTATCTTC	3240
AGGTAGATCT TTGAGGGCAT CTTTATTTGC AGGAATAGAT CCAAAGTTCT TACTATATTC	3300
TACTTGAATT TCTGATTGAC CAAACCAATC AATAAATCTT TTAGCTAACG CTTGTTTTTT	3360
ACTAGTGCTT AAAACCATAG TTTGTTCACT TACAAATGGT ACACCAATCT CAGGAGTCAT	3420
AACCTTGAAA ACAACATTTT GTTCTTTTGT TCCAATAAT GCACCAGAAC CCCACATCAT	3480
TCCATATTGT ATTGGATCTT CTTTGCTTAA CATCTTAACA ATTGAACTTT CTCCCTTTTG	3540
AAGAGTGTAT GCATTTTTC AATATTCTTT TGCTACTTCC CAACCTTTTT CGGAAACACC	3600
TAATTCACCT TTATCATCAA GGTATCGAAC TAAGATACTT GCTAGAATTG CCCGTCCTGT	3660
ACCTCCTTGA AGACCAGAAA TTGAATATTT ACCTTTTATC TTAATACCTA ATTCAGTCCA	3720
ATCTTTAGGC ATTTCTTTTA CATCAGGCGC CCCAATTAAA ACTAATGGTT GAACAATCAC	3780
AGGATTATAA TAATTATCTT TATCTGATAA AGATTGATCA ATTTTATCTA ACCATTTAGG	3840
CTTGTAATGT ACTAGTAATT TTTGATCTCT AATTTTATTT GAATCAACAG CACCAATTCC	3900
AAATACCATA TCTGCAACTG CATTATTCTT CTCAGCAATA ACACGGTCTG CTAATTGAGC	3960
GCCAGCGATA TCAACCATTT TTATATTAAA ACCAGCTTCT TTTGCTTTAG CAGTTAACCA	4020
ATCACCACGA CCAATTTGAGA CTGAGTTCGA ATAGATAACT AATCTTGAC TTTTATCAGC	4080

1135

```

TTTTTCTTCA GATGAAGAAG CAGTCGTAGA ATTTGAACCT CCAGAGCAAG CAGCAAGTGT      4140
AGTAAGAGCA ACTCCCGTTG CAAGTACAGT AGACCAAACCT TTCATTTTTT TCATGATAAG      4200
TTCTCCTTTT TTATTATTTT ATTTAAATTT TTCGTGATAT GGAACAAATT GTCTCATATC      4260
TTCAAATACA GTATAGTCAA TACGGTTTAC AGTAATAGTT GGAATCTTCT CTAATAAAAT      4320
TTCAGTTAAT TCTGCTCTGA CTTTAGTAAA CTCTTCTTCC TCCTCTTCGG TTAGAGGAAT      4380
CCGAAGATAC CCAATTGAAA TATGGAATTG ATATCTATCA TGATTAGGGA AACAAACACC      4440
TGCTTTTTCT GAGACATAAG TACGAATTTC TTCTAATCTC TTTGCAGAAG CTTTCATCTGC      4500
AGGTTCAACT AGTATGTTTT GTTTTCCCAT TTCAGTTATA CGCATATGAA TTTCTTCATC      4560
CAACAATGGA AAAATTTCAA GTTGTTTAGC AAAGTAATCA TGTATTTCTT GTAAAGGTGT      4620
ATCTAGAGGA AGATTACTGC TCCAAAACCTC gTCACGATT TTCATGGCAC AACAATTCAA      4680
TTACAGTCAT GTGAATAGAA TTCCTTGAGG TAAAGTAAA CTTATCGATA AATGGTAATT      4740
CTCTATAACG TGATTGAATA ATATCAACAA CTTCCATCAA ATCTTGTTTA GTATAAAGAT      4800
TTGCTACAAC TGTATTCCCA GGGAAATGAT TAAATTCCCC ATTCTCGG      4848

```

(2) INFORMATION FOR SEQ ID NO: 186:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3763 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

```

GTTATAAGCA ACACCTTCTT GCTTGCCATA AGTTGTGAAA TGGGTAGAAT CGATATCTAC      60
AATGAGTTGG TTTAGCTGGT GAAACTGTAA AAAGAATTCG ACCAATTCAA GCTTGAGGCA      120
TCGCAAACTA TGGACTGTTT CCTCGTCAGT TCTGGAAAGA AAACGGGATA AGGTTGGCTG      180
TGAAGCAAGC TGCCCTCCTT CCAATAATTT TGGAAAGTAG GCATCAGCTG ACAATTCTTT      240
ACAAGCATAG TCCGTTCCAT AACCTGTAA CAGTTGAAAG AGGAACTGGA CAAGGATATC      300
TGAATCCGAA TAACGACAGT AGCGGCGTTG GTCATTCTGT ACTAAATACT TAGAAATCCG      360
CTCTTTTAGT TTCAACTGGG AAAAAAGTTC CTGAAAAAAG ATAAGACCAC CATACTGGGT      420
TAAATGACCT CCATCGAAAG ATAGTTGGTA AAAAGACTTG TTTTGAAGT GATGATTTGG      480
TAAACTGTTT ATGTGAGTTT CCTTCTTTT TGTGTTTTTT TCTACACTTA TACCATAAAG      540
GGGAAACTCT TTTTGTCTA GTAAAAACA CCCATTGGGT GAAAAAGAA ACCATCCAGG      600

```

1136

ATCTAAGCTA AGGCAAGGAT TCTGGATGGT TTTTAGATTT GGGGTGAATA ATTGGGGATT	660
TAGGAGAAAT GATGGTATCT TCCAAATCAA AATCAACTTC ACTCCATAGT CTCAACTGAT	720
TGATTTTCCC ATCTTGATAG GTCACATCCT TGTCAAGGAT AAACGTAGTC AACACCTCAT	780
GTTGACCTTG ACACCTGATG TCATCTACCA AGAGCCAGAC ATCCTCTACC AACATGAGGA	840
TTTTTCTCCT GTGAAGATAA GGCAAATCAG GTTCTGCTGA CCAATAAGCC CCTCAATAT	900
AATGCACTCC CTCCTTTTCT TTATGGTGAC AAAACAGGGA GTGAGGATAG TATTCATATT	960
CCCAGGATCC CGTGATTCTT TCCGGAGCTT TCCCATCTAC AATGCAGGTC GAATGACTCC	1020
AAGCACTCTT TAAGAGATAA CGTTCATATA TCTCCCGATA AGAATAACGC CCAGCATCTA	1080
TGAAAATAGG TTGGCCTTGA TACTGTAAGC AAAAATATT CTCGTCACTA TGAATATGGG	1140
CACCTCCTAG CGGACCATTT TTGAAAAATA GATAACGATG TTCATCCTTA ATGCAGACAT	1200
GTCCAGAGTC TTCAAAGATC ATGGACTTAG GCTGCCAAGC TCTCTTTTCA AATTCCTGCA	1260
GTCGCTTGAC CTMTTCTCGC CCCAGGAACA AGAGGCTAAG CAAATCAACT TTAACATCCA	1320
GACCGTTAAG AAGGTCTTCC TGGTTCAAAA CCACAGCAGA CAGGCTCAAA ATTTCTGTCTG	1380
TTTCTGTAGA ATCGCTATCA CCAAAGCCA AAGTCCGTCC ATCTAAGCCT GTCATCATTT	1440
GAATATAGGT CGCCATCTTT TCCAGCAACT CTTGGTAACT ATCTTGCAAG TCTGGAAGCA	1500
AGAGACACAA ATCCAGCAAG GCTTTATAAA CCTCTACATG ATAGAGAATC GACTGTTCAA	1560
ACTGGCTTCC ATCTCTAAA ATCTGTGTCT CAATTTGCTG TTTCAACTCC TCTGAAGCAA	1620
AATGGTAAGC TTCTTCTAGA TCCATCTTAT CTGAAAAGAA ATGATAGATA GCAAGCATCG	1680
GAATTGTTTG TAAATCCCC CAGTTACTAA GGGTGACTT GCGCGGATAG TAGCTTTTCA	1740
TAAAGTCAAT CTGCTTTTCT AGACTGACCA AAATTTTCTC TAGTTCTTTC TCCTCTAGCA	1800
AGTCAAATTT CAAGAGGAGC AAGAGTAGTT TCAACCAAGT AAAGGAACGA ATACCCGTAT	1860
CCAAGGTTCT AGTCATCAAG GATTGAGGAG AAAATTCTCT CACCTGCTCA ATCCAATCAA	1920
ATAGAAAGAA CTGCACTTTT TGAATATAGT CCTTATCTCC TTCTACCAGA TACCCTATCA	1980
TAAACTGCAA GAGATATTCT TGTGATTGA GCATATAAGA CCATTCTGGA TCATCTTCAA	2040
ATACTTGATC CCATACCATC GGCTGGATTT GATGGATTTT TGAACAAGGC TCCATATCCC	2100
AAGGACTATC AAACATAAAA CGATTGTCCA TCAAGCGTTC AAGGGAACCTC TTGACTTTCT	2160
CATAGTCTTT TGAACAGTGC GACAAGATAT AATCACGACA TTGATTTCCA TCGACTCTTT	2220
CAAAAAATTG TCTTCTTCT TCTTTCATTA TCTATTACCA GAAAAAGAAC TACTTAAAAA	2280
GCAGTTCTTT TGTCTTTCCC ATTACACTTT CCTTTTCTAC ATGGATGACC ACACCTTTTG	2340
CAATCTGCAA GGAGACCAAG TCATCTTGGA TAGAAATGAT TTTTCCATGA ATTCCAGACA	2400



1137

ATAACAACAC TTCATCACCA AATGTTAAAG AAGCTAAATA CTCTTGTCTG TGCTCCATCT	2460
GTTTGCGAAG CAACTTTTGC TGACGAATAG AATGAAAGCT TGACAGTAAA AGGGGACTCA	2520
CTGCCAAGAC AATCACTATT CCATAAAACA ATGTTGTATC CATTAAAGCTA TAATCTTAAG	2580
CCAGCTTCCG ATAATTCCGA TGATAACTGT TAAATAACG AGTTTATATG TTGTCCATTT	2640
CTTTTCTTTG ATCAAGTAGT AAACTAAAAG TGTAATAGG GCTGGTAGAA GAGCTGGAGC	2700
AACCTTATCA AGCATTCCCT GAATACTTAC GATACTTTGT TTAGCGTCTG CTTTAACTTC	2760
CCCTGCAGCA AAGGTAATCG GCACCATAAT CTTAACAGAT GTCGCTGCCA AACCAGCAAT	2820
TACGTTACAC CGATAATATT GGCAATACGA GAAATCGTTG CCATCTGTTT GCTTAGTTTA	2880
TCAATCACAG TTGTTCTTAG TTTGTATCCA TACAGACCAG TTGACAATTT AATCGCTGTT	2940
AAAATCGTAT TCATCGCAAG GAAGAACAAG ATTGGACCGA CAACCAAGCC TTCTTGAGCA	3000
AACGAAGCTG CGATGGTTGA GAACAATGGA GCTAAACAGA ATTGAGAAAG AGAATCCCCA	3060
ATACCTGCCA ATGGTCCCAT CAAGGCCATC TTGATGCTAC GTGTTTCTTT TGCCGGACGG	3120
CCATTTTCCA ACATTACAAG ATGCAAGCTG GTAATAAAAG GCAGGAAGTG TGGGTTGGTA	3180
TTATAGAATT CACAGTTTTC TTCCAAGGCT TGGTAGAAAC CTTCTGATC CTCTCCATAG	3240
TGTTTTTTCA AAGCAGGATA CATCACATTG GCATATCCCA ACCCTTGATA GTTACTATAG	3300
TTAAATCCAT TTTGACAAAA GAATGCCCGC AAAGACGTTT TAAGATAATC ACGTTTGT	3360
AATTTGTTAG ATCCAGTCAT CGTGTGCTTC CTCCTCTACC ACATGATCCG CTGTTTTTGG	3420
CTTGTTATAA AATTCAATCA AAGCAAAGAT AGTACCTACA ATTGCAATAC CAATTGTTGG	3480
GATGTTTAGA TAAGCTGCAC AAACATATCC CAACAAGACA AAGGGAATCA ACTCTTTCTT	3540
AGCCATCACT GACAAGATCA TCGCAAAACC GATAGCTGGG AGCATTTTAC CAGCAACTGT	3600
CAAACCTGTA AGTAATACCG GTGGAATGTA GTCTACGAGT TTCAACAAGG TATCCATTGA	3660
AAGGGCACCA AGCAACCCAA GGTAAATCCA ATAAAGGCAA ACAACCAAAT TGTTGCATTT	3720
AGAGTGAAGT TAAATTTCTT CAAATTATGG TTTTCAAGT GCT	3763

(2) INFORMATION FOR SEQ ID NO: 187:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5053 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

1138					
CAATCTCTGA	GTATGTGCGG	TCAATACTAW	CAAAGGGAAT	YCCTGACGTC	AAGTAATGTT 60
CAATTGGmCT	ATAGGTAATG	GCAACCACTC	CATCAACTTT	ATTATGACGC	AACATCTCCA 120
GATAGTCTTG	CTCTCTATTT	GTACCATTGA	TAGAACATAA	GAGTAATTTG	TTATTTCTCT 180
TATAGACTTC	ATTTTCCACA	TGCATAGCAA	ATTCTGAAAA	GAAGGGATGC	CAGATACTTG 240
GTACAATGAT	TGCAATCGTT	TCTGTTGAT	TTTTTTTCAT	TCCTCTAGCG	TAGTAATCTG 300
GAATGTAATT	CAAAGTTTTA	ATCGCTTGTT	CCACTTTTTT	CAAAGTTACT	TCTTTAATGC 360
CTTTTTCTTT	ATTAATTACA	CGTGAAACAG	TTCCAACACT	AACTCCTGCT	TCTAAAGCAA 420
CATCTTTCAT	GGTAATTGAT	TTTCTTTGTT	CTACCATATT	ATCACCTCCT	TTCAATATAT 480
AGTATCATGC	AAATGCTTTT	TAAGCAACTA	TTTCTCAATC	ATTTTGGGCC	AGATCATTTA 540
TCCCATCATG	AATAAAATCA	CTCCAATTAG	CTTTTGAAAA	TACTTCAATT	TTCATGTGTA 600
AACATCTACA	TAAACAGGA	AAAGCCTTGG	TTTCATGGCT	TTTTTCGTAT	CTTCTATAAA 660
AAAAGCAAGA	GTTTGTAGATG	GCTATAAATC	TAGATGTACA	TTTTGCTTAA	ATGATTGAAG 720
GTCTTTTCTT	AACAAAAACA	CCCCCAAAT	TAGACTTTTT	CTGTCTAACT	TTTGAGGTAC 780
AGTTCAAACG	CGAAATAGCG	TTTTTTTGTT	ATTTTGGTT	ACTCATCTAA	TCGAATAAAC 840
ATCATGGCAT	TTAACAAGTA	TATGAGTGAG	ACCGTGTTTA	TATTATTTGA	ATAGATGAGT 900
CTCTTATTTT	CAATAGGAGG	AATAATAAAA	TTAGAAATAA	TGATATCATA	AGGTGAATCT 960
TCTAAAGATT	CCTTTGATAA	TTCTAATTCA	GTCCAAACTT	CCAGTTCAAA	ATTATTGCTA 1020
CAATAATAAG	AAAGTGCTC	TGCAACGAAT	TTTGCATGAT	ACTGATCAAA	ATTACTCATA 1080
ACTAAACCT	TTAGTTTAGG	CTGATTTTGT	AGCAAATTAA	TCACCAAATG	TTTGGTATGA 1140
GTGATGAAGG	TATAAGATAG	ATGATTTACC	ATCATTGAAC	TAGAACAAAC	CTCAAGAGTC 1200
TCTAAATAGT	GAGAAAGCTC	TTTTTTTATA	TCTGAAACAA	ATTTTGAAA	AATATTTTGA 1260
AAGTTCCTGA	TTGTATTCCC	TTTTTGATCA	AATAAAATAA	ACTCAGTAAA	CAACTCTTGA 1320
CGATACAGAT	GTGCGGTATT	ATGCAGATGC	CAATCAGAT	TATCCTTATT	CTCCATTTCa 1380
ATCTGATACT	TGACTGAAAT	CTGATCAATA	AAATCACTCA	ATAGATGGTA	AGATTTTTTCA 1440
ACATAACTAT	CCTTTTTTAC	GCATTTTCATA	AAGAGACTTT	CATCTATGAA	AAACATTTTT 1500
TGAAAGTAAG	ACACAAATAA	TTGGCAAACA	ACTTCTTCAT	CTAAAGAGAT	ATTGTATTCT 1560
GATTCAAAAC	TCTGAGCAAC	ACCTTCTATT	CCTTCTGCCT	GCATTAATAA	ATCCAAACTT 1620
TGGTCGTTAA	AAGAATCTTT	ATCTACTTCC	ATAAAATGAC	CAAACCTTTAT	TCTATATAGG 1680
TTTCGTAAC	GGAGCAACTT	TAGCATTTCTA	TGCGTTGACA	AATTCATTGG	AAAGCTTGTT 1740
TCCTTATAAA	CCAATTCTAA	CAATTGAGAT	AGTGGCTCTG	ATGAAAAATT	TTCAAATGGC 1800

1139

CATTCTAGGA AATAATATTT TTCTGAAAAA TATTGTGCAA AAAAGTAACG AATGTCTCTC	1860
TCATTTCCAA TGATTTGAAC AGGGGTCAGA CTAACCTCAA ATTGAAATTG CCTTTTAATC	1920
ACTTTATTGA TTTGGCTAAT AATACGATAG AGCGAAGATG AACTGATATA AAATTCTTTA	1980
CAAATACTCT CAGCTTGACA ACCTTCATTA AAGAAGATGA ATTCTAAAAT CGAAAAATGA	2040
GTTGAATGTT TAAAGAAATG ATGGTAAACC ATTTCAATAT CACTATCATC GGTATTAATA	2100
ATGCGTATAC CATTAGTAGA AGAATGAAAA ATCAAGTCAG GAAAAGCAGA TTTAACATGG	2160
GATAGATCAT CTTGACTGC ACGTTCTGTA CAATTTAATA ACTCTGCTAG TTCAGAACGA	2220
TGAAACCAAC GTTTATGTTT AAATAATAAT TCTAATAATT CTAATTCCTT ATGACTTTTT	2280
TTAGATAATA AATCTCTCAT GAATATCTTT CTCTCTTTAT AAATTATCGG ATTAAACCTC	2340
TTGCAATTAT ACCACAAAGA ATAGGTATAG CATGATATAA CGACTTTTCC TAAAACTTTT	2400
TATTTCTGAT AATAACACTA CGGAGACAAT ATATAAACAA TTTTCTTATT TTACCGTCTA	2460
TTGAGGGCGT GAATACAGAA TCAAAATCAA GTCTAAAGAT TATATTTTAA ATTTTAAAAA	2520
TTATATAATA GCAACAATTA AAGAATTTGA TTTTAAAAA TTATATAATA ATAACAATCG	2580
AAATAATTGA CTTTTCTATA TTAAAGTTAT ATAATAGTAA TAATCAAAGA AATTGATTTT	2640
TTGATATTAA AATAAAAAAG GAGGGTAGGC AGTGTGTGTA TCAATTATTG CTGGAGGTCT	2700
TATTTGCTCTC TTGGCAGGTA AAATCACTAA AAAAGTAGTT CTATGGGAAT CATCGCAAAT	2760
GTATTCGCTG GTTTAGTCGG GGCATATGCA GGACAATCTC TTTTAGGTAG TTGGGGTCCA	2820
GCAATCGCTG GAATGGCTTT GCTCCCATCT ATTGTAGGTG CAGCGATTGT GATTACTGTA	2880
GTGTCAATCT TTACAGGTAG AAAGTAACT TTTCCCCAGT AAAGTTAGCA AACTATTTTT	2940
AAATCAATGA CGGGAAAAAT AGTTTAAATG TTAAATCGAA AGGATTGTAT ATGTCAAAG	3000
CAAAGAAAAT ATGTTTCATT ATTTTCTGTA TTTTAATCTT GACAATTTTC CTTCTGTGTT	3060
TGATAGATTA TCATCAAGTT AGTGATCTAG GTATTCATCT ACTTAGCTGG AGACAGAACT	3120
CCGTAGTTGA ATTCTATCTT GCTAGATATG TCTTTTGGGG GACAGTGGTT CTATCAACTT	3180
TAGTTTTATT ATCCATTTTA GTTGTGATGT TTTATCCTAA ACGTTACTTG GAAATCCAAC	3240
TTGAAACTAA AAACGATACA TTAAATTTAA AGAATTCGGC AATCGAAGGT TTTGTTAGAA	3300
GTTTGGTGAG TGATCATAGA TTGATCAAGA ACCCAACTGT TCATGTAAAT TTACGAAAAA	3360
ATAAATGTTT CGTTCATGTA GAAGGTAAAA TTCTTCCTTC AGACAACATC GCTGACAGAT	3420
GCCAAATAAT TCAAAATGAA ATAACATAAT GATTGAAGCA GTTTTTTGGT ATTGAGCGTC	3480
AAGTAAACT TGAAGTTGCA GTAAAAAATT ACCAACCAAA ACCTCAAAAC AAAAAGACTG	3540

1140

TTAGTCGTGT GAAGTAAGGA AGTAAAAAAT GGAATGGCTT AAACAATATC GATATCCAAT	3600
TATCGCTGGT CTCATAGGCG TATTTCTGGC TTGTTTGATT GTCTCCTTTG GCTTCTTCAA	3660
AAACAATATTT GTATTGATTT TAGGAGCACT GGGAGTTGCA GCTGGATTAT ATATCGAAAA	3720
AAACTATATA GATAAATAAA AAAATAAAAA TTACTAATTT AATTAAAGGA GTTTCATATG	3780
TCAAACGAAA AAAACACAAA CACTAACGTA GAAAAGAAAG ATGCTACTGT TGTAGCTCAC	3840
GAAATCAAAG GGGAACTTAC TTACGAAGAT AAAGTTATCC AAAAAATCAT TGGTCTTTCA	3900
CTAGAAAACG TTTCAGGTCT TTTGGGAATC GATGGTGGTT TCTTCTCAA TCTTAAAGAA	3960
AAAATCGTTA ACAGCGATGA CGTAACAAGT GGTGTTAACG TAGAAGTTGG TAAACACAA	4020
GTTCAGTTG ACTTAAACGT TATTGTTGAG TACCAAAAAA ATGTTCCAGC TTTATATTCA	4080
GAAATCAGAG AAATCGTATC TTCAGAAGTT GCTAAAATGA CTGACTTGG AATTGTTGAA	4140
ATCAACGTAA ACGTTGTCGA CATCAAACT AAAGAACAGC ATGAAGCAGA CTCAGTAAGC	4200
CTTCAAGATC GCGTATCTGA CGTTGCTGAA TCAACAGGAG AATTCAC TTC AGAACAATTC	4260
GAAAAAGCTA AATCTGGTCT TGGATCTGGT TTCTCACTG TTCAAGAAAA AGTTAGCGAA	4320
GGTGTAGAAG CTGTTAAAGG TGCAGCAAAT GGTGTAGTAT CTCACGAAAA CACTCGTGTA	4380
AACTAAGATA AAATAAATAT AACAGGAGAA ATTATCATGT CAGTAGAAGA AAAATTAAAT	4440
CAAGCTAAAG GTTCTATTAA AGAAGGTGTT GGGAAAGCCA TCGGTGATGA AAAAATGGAA	4500
AAAGAAGGTG CAGCTGAAAA AGTTGTTTCT AAAGTAAAG AAGTTGCCGA AGACGCTAAA	4560
GACGCTGTAG AAGGTGCTGT AGAAGGTGTT AAAACATGT TGAGTGCGCA CGATAAATAA	4620
GGTAAAAAGT TACTTTATCT TTTTAGTAAT ATTAGTCAAA AGAGTCTGAG TCAAGATGAT	4680
TCTCAGAAAA CAAAAGCTA GAGATTCCCA ATTGCGGAAC TCTAGCTTTT TAATTTTGCC	4740
TCTTTCTCTT ATTATATTTC AGCAGGTGTG TGGCCATGAG TACGAATCCC ATGTCAATTC	4800
TCACTTGACG CTTACCTCTC AGATGACATC TCTTATAACC CAAACAAACC TTTATCTGCC	4860
CAAAGACAGA TTTCATATCA ATCTTACGTT TAGCGAAAAT TTGTCTACCC TTGGAAGATA	4920
AAAGTGCTG ATATTCTTTA GTTTTAAAC ACTGGTAACG TTCATTCTATA TACAGTCTCT	4980
TTTGAGGGGC TGATTCAGGT TCATAATCGC AGTCAACATT GATTTCAAGG CTGTTTGCTT	5040
TCTATCTCCC CGG	5053

(2) INFORMATION FOR SEQ ID NO: 188:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6492 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1141

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

AATCTCTTT TTTCCAACAA AATGTATGAC CTGCACTTGA ATACTTCTCA TTGTTTGTAC	60
ATTCATCTAC TTTCATATAA TCTTTTACAA AATCATAATA TGACATAACA CACTATCCCT	120
TTTAGACAAT ATTCCAATTA GCCTTATTAA TTCAAACTA TTGTATTAGT AATTATAACA	180
GATGTATAAT AGAAAAGCAA TGATAGATAT TATCAATTAA GCGAATTTAT ATCTAAAAGG	240
GATATTAAAG AAAGGAGATA TGCTTATGAA GATTTACAAA AAACATTTTG CTTATGTCCA	300
AGATAAGAAA TATCTTGGGG TTTTGGCCAT AATTTTTTCT GCTATATCTG CTGCACTTAC	360
AGTATATGGA TATTATTTAA TCTACAAAT TCTAGATAAG TTAATAATTA ATTCAAACCT	420
ATCCGGTGCA GAGAGTATAG CATTAAATC TGTTATTACA CTAACAAGTG GAGCGATATT	480
TTATTTTGTG TCAGGAATGT TTTCACATAT CTTGGGATTC AGGCTTGAAA CAAATTTAAG	540
AAAAAGGGAA TCGATGGTCT GGAAAAAGCA AGTTTTAGGT TCTTTGACTT AAATCCATCT	600
GGTCAAATAA GAAAGATTAT AGATGACAAT GCTGCACAAA CTCATCAGGT GGTAGCACAC	660
ATGATCCCG ATAGTTCTCA GGCAATAATC ACACCCGTAC TTGTACTTGC ACTTGGCTTT	720
ATAGTAAGTA TAAGAGTTGG CATAATTTTG CTTGCTCTTA CTATAATTGG TGGCTTAATT	780
TTAGGGGCAA TGATGGGCGA GCAAGAATTT ATGAAGATAT ACCAAGAATC CCTATCTAAA	840
CTAAGTGCTG AAACGTGTGA GTACGTGAGA GGAATGCAAG TTGTAAAAAT ATTTAAAGCA	900
AATGTAGAGT CTTTTAAAAG CTTTTATAAG GCGATAAAAG ATTACTCAA GTATGCTTAT	960
GATTATCCC TATCTTGTA AAGGCCTTAT GTTTTGTATC AATGGTTATT TTTGGACTG	1020
ATTGCAATTT TAATTATTCC TATAGTTTAT TTTATGACTA GCTTAGCTAG CGCAAAGGTG	1080
ATTTTACTAG AGCTTATCAT GATTTTATTT TTATCAGGAG TTCTCTTGT TTCATTCATG	1140
AGAATGATGT GCACTCCATG TATATTTCTC AAGGAAATTA TGCAGTAGAT ACTTTAGAGG	1200
CGCTTTACGA AGATATGCAA AAAGACAAAT TAGTGCATGG TAATGTCAAT AATTTTAAAA	1260
ACTATAATAT AGAATTTGAG AATGTTAGCT TTGCTTATAA TGATAAAGCT GTCATTGAAA	1320
ATTTATCCTT TAATTTAGAA GAAGGAAAGT CCTACGCACT TGTGGTTCA TCTGGATCAG	1380
GCAAATCAAC AGTAGCAAAA CTTATATCAG GTTTTACAA TGTTAATAAA GGAAGCATAA	1440
AGATAGGCGG GATAGCAATA AGTGAATATT CTGACGAAGC CTTAATTAAA GCCATTTCTT	1500
TTGTTTTTCA AGATTCAAAA TTATTCAAGA AGAGCATTTA TGATAATGTA GCGTTAGCTA	1560
ATAAAGATGC GACGAAAGAT GACGTTATGA GAGCCTTAAA ATTAGCAGGA TCGATTTAA	1620

1142

TATTAGACAA ATTCCCAGAA AGAGAAAATA CAATCATAGG CTCAAAAGGT GTTTATTTAT	1680
CCGGTGGAGA AAAACAAAGA ATTGCAATTG CTAGAGCAAT TTTAAAGGAT TCCAAAATTA	1740
TTATTATGGA TGAAGCATCA GCATCTATTG ACCCAGATAA CGAGTTTGAA TTGCAAAAAG	1800
CTTTTAAAAA TCTTATGAAG GATAAAACAG TTATCATGAT TGCACACAGG CTATCTACAA	1860
TTAAAGACCT TGATGAAATT ATTGTCATGG ATAGTGGAAA AATTATAGAA AGAGGGTCTG	1920
ACAAAGAATT AATGTCAAAA GATACAAGGT ATAAGAGCCT GCAAGAGATG TTTAACAGTG	1980
CGAATGAATG GAGGGTTTCA AATGAAAGAG TTTTATAAAA AAAGATTTGC TCTTACAGAT	2040
GGAGGAGCAA GAAATTTAAG TAAAGCAACA CTGGCTTCAT TTTTCGTTTA TTGTATAAAC	2100
ATGCTTCCTG CCATATTACT TATGATTTTT GCTCAGGAAG TTTTGAAAA TATGGGCAAA	2160
AGCAATGGCT TTTATATAGT ATTCTCAGTT TTGATTTTGA TAGCAATGTA TATTTTGCTT	2220
TCTATCGAAT ACGATAAATT ATATAACACA ACCTATCAAG AAAGTGCAGA TTTAAGAATA	2280
AGGACAGCGG AGAATTTATC AAAATTACCT CTATCTTACT TTTCTAAACA TGACATTTCC	2340
GACATTTTAC AAACAATCAT GGCTGATATT GAAGGCATAG AGCATGCAAT GAGCCACTCA	2400
ATACCAAAGG TGGCGGCAT GGTACTGTTT TTCCCATTA TATCTGTAAT GATGCTAGCG	2460
GGCAATGTCA AGATGGGTTT AGCTGTAATT ATTCCATCTA TTTTAAGCTT TATATTTATA	2520
CCTTTATCTA AAAAATATCA GGTTAATGGA CAGAATAGAT ATTATGATGT CTTAAGAAAA	2580
AACTCAGAAA GCTTTCAAGA AAATATCGAA ATGCAAATGG AGATTAAAGC ATATAATTTA	2640
TCAAGGATA TTAAAGATGA CTTATATAAA AAAATGGAAG ATAGTGAGAA AGTACACTTA	2700
AAGGCGGAAG TAACTACAAT TTTAACTTTG TCTATATCTT CAATATTTAG CTTTATATCT	2760
CTTGCTGTTG TGATATTTGT CGGCGTAAAT CTAATTATTA ATAAAGAGAT AAATCTCTCT	2820
TACCTTATAG GATATTTACT AGCTGCTATG AAGATAACAG ACTCTTTAGA TGCATCTAAA	2880
GAGGGCTTGA TGGAAATATT TTATTTATCG CCCAAAATAG AAAGATTAAA AGAAATTCAA	2940
AATCAAGATT TACAAGAAGG CGATGACTAT AGCTTAAAAA AATTGATAT TGATCTAAAA	3000
GATGTTGAGT TTGCCTACAA TAAAGACGCA AAAGTTTAA ATGGTGTAAG TTTTAAAGCT	3060
AAGCAGGGAG AGGTCACTGC TTTGGTAGGT GCAAGTGGCT GCGGTAAAAC AACTATCTTG	3120
AAACTTATAT CAAGACTTTA TGATTATGAC AAGGGACAAA TCTTAATCGA TGGCAAAGAT	3180
ATAAAGGAAA TATCAACAGA ATCCCTTTTT GATAAGGTGT CTATTGTTTT CCAAGATGTG	3240
GTTCTCTTTA ATCAAAGCGT TATGGAAAAT ATTAGAATCG GTAAGCAAGA TGCAAGTGAC	3300
GAAGAGGTTA AAAGAGCAGC AAAACTTGCA AATTGCACAG ATTTTATAGA AAAAATGGAT	3360
AAAGGTTTCG ATACAGTTAT TGGTGAAAAC GGAGCTGAGC TATCAGGAGG AGAAAGACAA	3420

1143

AGATTATCAA TAGCCAGAGC CTTCTTAAAA GATGCGCCGA TATTGATCTT AGATGAGATA	3480
ACAGCAAGCC TTGATGTTAA CAACGAGAAA AAGATTCAAG AGTCTTTAAA TAATTTAGTT	3540
AAAGATAAAA CTGTTGTAAT CATTTACAT AGAATGAAAT CCATAGAAAA TGCAGACAAG	3600
ATAGTAGTTC TTCAAAACGG AAGAGTAGAA AGCGAAGGTA AGCATGAAGA GCTTTTACAA	3660
AAATCAAAAA TTTACAAAAA TTTAATAGAA AAGACAAAAA TGGCAGAAGA ATTTATTTAT	3720
TAGGAGGACT ACAATGGATA ATAAAAAATT AAAAGTAAAA GATTTAGTAA GCATCGGTGT	3780
TTTTGGCGTA ATTTATTTTG CCTTCATGTT TGGAGTTGGT ATGATGGGCT TGATTCCAAT	3840
ATTGTTCTTA ATATACCCGA CAGTATTAGC CATAGTTGCA GGAAGTGTG TTATGTTATT	3900
TATGGCTAAG GTTCAAAAGC CATGGGCACT ATTTATATTT GGTATGATAT CACCACTTGT	3960
GATGTTTGCA GCTGGTCATA CCTACGTAGT TGTGGTTTTA TCACTTATAG TAATGATAAT	4020
AGCAGAATTA ATTAGAAAGA TTGGTAATTA TAATTCATTT AAATACAATA TGCTTTCTTA	4080
TGCAATCTTC AGCACATGGA TATGTAGCTC TTTAATGCAA ATGCTTTTAG CAAAAGAAAA	4140
ATATATGGAG TGGTCTTTGA TGAATATGGG AAAAGATTAT GTTGATGTAT TAGAAAAGTT	4200
AATAACTTAT CCTCACATGG CTTTAGTAGC CTTAGGTGCT TTCTTAGGAG GAATTCCTGG	4260
AGCATATATA GGCAAGGCTC TATTGAAAAA ACACTTTTCA AATGGATTAT ATTGTGTGGG	4320
ATACTTTACT CCTTGCCTAA TTTTATGGTG CTATCTGAAT TAAACCTAT AGTTAAGATG	4380
TTTTTGAGTA TACCTATTGT TATTAGAATG TTTATTTTAC CATTTATGGC AGCAAGCTTT	4440
ATGATAAAGA CCTCGGATGT AGGCGCAATA ATTTATCGA TGGATAAGCT TAAGATTTC	4500
AAGAATGTAT CCATACCTAT TCGCGTTATG TTTAGATTCT TCCCATCTT TAAGGAGGAG	4560
AAGAAAAACA TCAAAATGGC TATGAGAGTA AGAGGGATAA ATTTTAAAAA CCCAGTCAAA	4620
TATCTTGAAT ATGTTTCTGT GCCACTACTC ATTATATCAT CTAATATATC AGATGACATT	4680
GCAAAAGCGG CAGAAACAAA GGCAATAGAA AATCCAATTG CCAAGACCAG ATACATTCCG	4740
GTAAAGATAC AGCTAATTGA TTTTGTTTAT GTTTTAGCGG TTGCTGGACT TATTGTGGGA	4800
GGCTTAATAT GGTGAAATA AAAAATTTAA GTCTTGATTA TGGTGAAGAG CATATATTAG	4860
ATGATATATC ACTATCCATA GCCGAGGGAG AGTGCCTGCT ATTTACAGGA AAAAGTGGAA	4920
ATGGTAAGTC ATCTTTAATA AATTCAATCA ATGGACTAGC TGTAAGGTAT GATAACGCAA	4980
AGACAAAGGG CGAAATAATT ATTGATGGTA AGAATATAAA AAATTTGGAA CTTTATCAAA	5040
TCTCAATGCT TGTTTCAACT GTTTTTCAAA ATCCTAAGAC ATATTTTTTT AATGTCAATA	5100
CGACATTAGA ATTATTATTT TATTTGAAAA ATATCGGTCT TGCAAGAGAA GAGATGGACA	5160

1144

GGCGTTTGAA GGATATACTT GAGATATTCC CGATAAAAAA TCTTTTGAAC AGAAATATAT	5220
TTAATCTATC CGGCGGTGAA AAACAAATTC TTTGCATTGC AGCTTCTTAT ATAGCAGGTA	5280
CAAAGATTAT AGTTATGGAT GAGCCTTCAT CGAATTTAGA TATTAAAAGC ATAAGTGTTT	5340
TGGCAAAGAT GCTAAAGATA TTAAAGAGA AAGGCATAAG CATAATTGTT GCAGAGCATA	5400
GAATTTATTA TTTGATGGAC ATAGTTGACC GTGTATTTTT AATAGATAAA GGAAAGCTTA	5460
AAAAAAGCTTA TACTAGAAGT GAATTTTAA AGCTAGATAA AAATGAATTA AATGCTTTAA	5520
GTTTAAGAGA TAAAGAATTA AGTAAATTAA AAGTTCCTTA TTTAAAAGAA GGTGGAGAGT	5580
ATCAGATAAA AAATCTTAGT TACAAATTTA CTGATGATGA GTGTTTAAGC TTAAAGATA	5640
TTTCGTTCAA GCTTGGGAAA ATTTATGGCA TAATAGGATC CAACGGACGA GGAAATCAA	5700
CGCTTTTAAAG ATGTTTAATA GGTCTTGAGA AAAAATCAAA AGAAGAAATT TATTTTAAAG	5760
GAGAGAAGCT ATCTAAAAAA GAAAGACTCA AAAACTCTTC ACTTGTTATG CAAGATGTAA	5820
ATCATCAATT ATTCACAGAT GAAGTATTCA ACGAGCTTAG ATTAGGAGTA AAGAATTTTG	5880
ATGAAGAAAA GGCGAAAATC ATTTTAAACC CCAATTATTC ACCCCAAATC TAAAAACCAT	5940
CCAGAATCCT TGCCTTAGCT TAGATCCTGG ATGGTTTCTT TTTTCACCCA ATGGGTGTTT	6000
TTTACTAGAC AAAAAAGAGT TTCCCCTTTA TGGTATAAGT GTAGAAAAAA ACACAAAAAG	6060
AAAGGAAACT CACATGAACA GTTTACCAAA TCATCACTTC CAAAACAAGT CTTTTTACCA	6120
ACTATCTTTC GATGGAGGTC ATTTAACCCA GTATGGTGGT CTTATCTTTT TTCAGGAACT	6180
TTTTTCCCG TGAAACTAA AAGAGCGGAT TTCTAAGTAT TTAGTAACGA ATGACCAACG	6240
CCGCTACTGT CGTTATTCGG ATTCAGATAT CCTTGTCAG TTCCTCTTTC AACTGTTAAC	6300
AGGTTATGGA ACGGACTATG CTTGTAAAGA ATTGTCAGCT GATGCCTACT TTCCAAAATT	6360
GTTGGAAGGA GGGCAGCTTG TTCACAGCCA ACCTTATCCC GTTTTCTTTC CAGAACTGAC	6420
GAGGAAACAG TCCATAGTTT GCGATGCCTC AACCTTGAAT TGGTCGAATT CTTTTTACAT	6480
GTTCAACGAGC TG	6492

(2) INFORMATION FOR SEQ ID NO: 189:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 7174 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

AACTGAAGGT AAAGGCTTCG ACGCAGAACG TGACGCTGCC CAAGCTGCCC TTGATGACCT	60
---	----



1145

TAAGAAAGCT CAAGAAGACA ACAACTTGGG CGACATGAAA AAAAACTTG AAGCATTGAA	120
CGAAAAAGCT CAAGGACTTG CTGTTAAACT CTACGAACAA GCCGCAGCAG CGCAACAAGC	180
TCAAGAAGGA GCAGAAGGCG CACAAGCAAC AGGGAACGCA GGCGATGACG TCGTAGACGG	240
AGAGTTTACG GAAAAGTAAG ATGAGTGTAT TGGATGAAGA GTATCTAAAA AATACACGAA	300
AAGTTTATAA TGATTTTGT AATCAAGCTG ATAACATAG AACATCAAAA GATTTTATTG	360
ATAATATTCC AATAGAATAT TTAGCTAGAT ATAGAGAATT ATATTAGCTG AACATGATAG	420
TTGTATCAAA AATGATGAAG CGGTAAGGAA TTTTGTACC TCAGTATTGT TGTCTGCATT	480
TGTATCGGCG ATGGTACCAG CTATGATATC ATTAGAAATA CAAACATATA AATTTGTAAT	540
ACCGTTCATA ATTGGTATGA TTTGGACAGT AGTTGTATTT CTTATGATCA ATTGGAATTA	600
TATAGGCAAA TACTAAGAAG AGACAAAAAT ATATAAATAT TTCTGTACTT ATAGGATATT	660
TAAATCAAA ATAAAGTTAA TTTACTTATT TGCAGAGGTT GCAACCCAGC CTCTGTTTTT	720
CGATAAAAAG GGACGGAATC TCATTTGTTT GGGTTTTGTC TCATCAATAG AAAGGAACAA	780
AGAGTGTTTG TAACTGAACA CGGGTTTCAG AATTTCTTAC TAAATATAAA AGAAAGGAAT	840
TGAACCCGAC CTAAATGGTG GTTCGATTCA GAACATCAAT AGAAAGGAAT AAGGGTGTTT	900
GTAACCTGAC ACGGGCTATG GACTGTGCCA AAAAGATAGT TTTTCTAGG ACCTAAGCGT	960
CCGTCGTCAA AACTCCTAGA TGGCTGTGTC CGTTTGACGC CCTTTGTATC TTGAATTATG	1020
AACAATACTG AATTTTATGA TCGTCTGGGG GTATCCAAAA ACGCTTCGGC AGACGAAATC	1080
AAAAAGGCTT ATCGTAAGCT TTCCAAAAAA TATCACCAG ATATCAACAA GGAGCCTGGT	1140
GCTGAGGACA AGTACAAGGA AGTTCAAGAA GCCTATGAGA CTTTGAGTGA CGACCAAAAA	1200
CGTGCTGCCT ATGACCAGTA TGGTGCTGCA GGCGCCAATG GTGGTTTTGG TGGAGCTGGT	1260
GGTTTCGGCG GTTCAATGG GGCAGGTGGC TTCGGTGGTT TTGAGGATAT TTTCTCAAGT	1320
TTCTTCGGCG GAGGCGGTTT TTCGCGCAAT CCAAACGCTC CTCGCCAAGG AGATGATCTC	1380
CAGTATCGTG TCAATTTGAC CTTTGAAGAA GCTATCTTCG GAACTGAGAA GGAAGTTAAG	1440
TATCATCGTG AAGCTGGCTG TCGTACATGT AATGGATCTG GTGCTAAGCC AGGGACAAGT	1500
CCAGTCACTT GTGGACGCTG TCATGGCGCT GGTGTCATTA ACGTCGATAC GCAGACTCCT	1560
CTTGGTATGA TGCGTCGCCA AGTAACCTGT GATGTCTGTC ACGGTCGAGG AAAAGAAATC	1620
AAATATCCAT GTACAACCTG TCATGGAACA GGTCATGAGA AACAAGCTCA TAGCGTACAT	1680
GTGAAATCC CTGCTGGTGT GGAACAGGT CAACAAATTC GCCTCGCTGG TCAAGGTGAA	1740
GCAGGCTTTA ACGGTGGACC TTATGGTGAC TTGTATGTAG TAGTTTCTGT GGAAGCTAGC	1800

		1146	
GACAAGTTTG AACGTGAAGG AACGACTATC TTCTACAATC TCAACCTCAA CTTTGTCCAA	1860		
GCGGCTCTTG GTGATACAGT AGATATTCCA ACTGTTACAG GTGATGTTGA ATTGGTTATT	1920		
CCAGAGGGAA CTCAGACTGG TAAGAAGTTC CGCCTACGTA GTAAGGGGGC ACCGAGCCTT	1980		
CGTGGCGGTG CAGTTGGTGA CCAATACGTT ACTGTTAATG TCGTAACACC GACAGGCTTG	2040		
AACGACCGCC AAAAAGTAGC CTTGAAAGAA TTCGCCGCTG CTGGTGACTT GAAAGTAAAT	2100		
CCAAAGAAAA AAGGCTTCTT TGACCATATT AAAGATGCCT TTGATGGAGA ATAATACTCT	2160		
TCGAAAATCT CTTCAAACCA CGTCAGCGTT GCCTTGCCGT ATATATGTGA CTGACTTCGT	2220		
CAGTCGTATC TACAACCTCA AAACAGTGTT TTGAGCAGCC CGTGGCTAGT TTCCTAGTTT	2280		
GCTTTTACT TTATAGATTT TTTAAGACTT TCCTAAGTAA TGACGGACGG TAGTGACCTC	2340		
CTTCGAAGTT CCATACCTAA ACTTTGAACC TAAGTTTTAA AGTTTCCGGA CAGCTGAAAC	2400		
CAAGCTGTTT CAGGTGTTTT CATTACGGCA GAAAGTCTTC GATTTAGTTG TGAAATGGTG	2460		
AATGATACTC TTCAAAAATT TCTCAAACC ACGTCAGCGT CGGCTTGTC TGGGTATGGT	2520		
TACTGACTTC GTCAGTTCTA TCCACAACCT CAAAACAGTG TTTGAGCTGA CTTGCTCAGT	2580		
TCTATCCACA ACCTTAAAAC GGTGTTTTGA GCAGTCTGTG CCTAGCTTTC TAGTTTGCTT	2640		
TTTGATTTTT ATTGAGTATG AATTACCTAA ATTATGATGC ATAGTTGATG GGATATATAT	2700		
AATAGATTGA AATAGAATAT GAACAAATTG ATAAGAGGAT TTAAAGTAA TCTCTAACAA	2760		
TGCTTTAGAA ACTATGGTGT GCTATTCTAA ATTCAATTCA CTATAACTTG TTTACGTTTT	2820		
AAAAAAGAGC CGTCGGGCTC TTTTACTTA TCTTCAGTTC CCTGCATTTT TTTATCACA	2880		
GCTAGTCTAG TCTGGATATC CTTTTCCAAG ACCTTAACT TGTAAGTCAA GTCTTCTTGG	2940		
TATTCCTTGA TAAGTTCTTT TTGCTGGTTA ATGATTTGCA GGCTGTTTTG GATAATATCC	3000		
ACATCGTCCT TGATAGCTTG AACGCGGTCA GTGGTATTCA AGACTTCATC TGTGATGGTT	3060		
TGGCGATTTT TTGTAACCAG ATAACCTCCG GCTGCAGCTC CTGCAAATAG CAGTAGGTTG	3120		
GATAATTTCA TAGCAACTCC TTAAGCGTTT TTGATGGTTT CAGCGACTTG AGCAAGTTTG	3180		
TCAAAGCTG GTTCGTGGGC GATAAAATCA ATCTTGAGGT CATCGTCAGC ACTGTAGCGA	3240		
GGCACAAGGT GAACGTGAGT ATGAAAACT GTTTGACCAG CGACTTCTTC ACAGTTGGAA	3300		
ATGATATTCA TACCAGCAGC CTTAGTGACT TTCATGACTT TTTGAGCTAC TTTTGGTACT	3360		
TGGGCAAAGA GTTGGCTGGC GCTCGTAGCA TCCATCTCCA AAAGATTGCG ATAGTGTCTT	3420		
TTTGGCACGA CCAAGGTGTG TCCTAGTGTT ACTTGAGAGA TATCAAGAAA GGCAAGGACC	3480		
TGCTCATCTT CATATACTTT TGAAGCAGGA ATTTCCCCTG CGATGATTTT AAAAAAATG	3540		
CAATCTGACA TAAATCTAC CTCTACTGTA CTGAATTTG ATATAATATA GCTACATTAT	3600		

1147

ACCAGATTTG	GAGAAAATAT	GTTAGAAATT	AAAAACCTGA	CAGGTGGCTA	TGTTCAATGTT	3660
CCTGTTTTGA	AAGATGTGTC	CTTTACTGTT	GAAAGTGGGC	AGTTGGTCGG	TTTGATTGGT	3720
CTCAATGGTG	CTGGGAAATC	AACGACGATC	AATGAGATTA	TCGGTCTGTT	GGCACCTTAT	3780
AGTGGCTCCA	TCAATATCAA	TGGCCTGACT	CTGCAAGGAG	ATGCGACTAG	CTACCGCAAG	3840
CAGATTGGCT	ACATTCTCTGA	GACGCCTAGT	CTGTATGAGG	AATTGACCCT	CAGAGAGCAT	3900
ATCGAAACGG	TTGCTATGGC	TTACGGTATT	GAGCAAAAAG	TGGCTTTCGA	ACGAGTAGAG	3960
CCCTTGTTAA	AAATGTTCCG	TTTGGAACAG	AAATTAGACT	GGTTCCTGT	TCATTTTCA	4020
AAAGGGATGA	AGCAGAAGGT	CATGATTATC	TGTGCTTTTG	TGGTGGATCC	AAGTCTTTTC	4080
ATCGTGGATG	AGCCTTTCCT	TGGTCTTGAT	CCGCTGGCTA	TTTCTGATTT	GATTACGCTT	4140
TTGGAAGTGG	AGAAGCAAAA	GGGCAAGTCT	ATTCTCATGA	GTACCCACGT	GCTGGATTCTG	4200
GCGGAGAAGA	TGTGTGATGC	CTTTGTCATT	CTTCACAAGG	GAGAGGTGCG	TTCCAAAGGC	4260
AATCTCCTGC	AACTACGTGA	AGCCTTTGAT	ATGCCTGAGG	CTAGTTTGAA	TGATATTTAC	4320
TTGGCTCTGA	CCAAAGAGGA	GGATCTATGA	AAGACTTGTT	TTTAAAGAGA	AAGCAGGCCT	4380
TTCGTAAGGA	GTGTCTTGGT	TATCTGCGCT	ATGTGCTCAA	TGACCACTTT	GTCTTGTTCC	4440
TGCTTGTCTT	GTTGGGCTTT	CTAGCCTACC	AGTACAGTCA	ACTCTTACAA	CATTTTCTCTG	4500
AAAAATCATG	GCCTATCCTT	TTGTTTGTAG	GAATTACGTC	TGTTTTACTT	TTACTTTGGG	4560
GAGGAACTGC	CACCTATATG	GAGGCTCCAG	ACAAGCTCTT	TCTCTTAGTT	GGAGAAGAGG	4620
AAATTAAGCT	CCATCTCAAG	CGTCAAACTG	GCATTTCCCT	AGTCTTTTGG	CTCTTTGTAC	4680
AGACCCCTTT	CTTGCTGTGA	TTTGCGCCTT	TATTTTTAGC	AATGGGTAT	GGCTTGCCAG	4740
TTTTTCTGCT	CTATGTGCTT	TTATTGGGGG	TAGGAAAATA	TTCCACTTT	TGTCAAAAGG	4800
CCAGCAAAAT	TTTCACTGAA	ACTGGACTGG	ACTGGGACTA	TGTTATTTCT	CAAGAAAGCA	4860
AGCGTAAGCA	AGTCTTGCTT	CGTTTCTTTG	CCCTCTTTAC	GCAGGTCAAG	GGAATTTCAA	4920
ACAGCGTTAA	GCGTCGTGCC	TATCTGGACT	TTATTTTAAA	GGCTGTTTCA	AAGGTGCCTG	4980
GGAAGATTTG	GCAAAATCTC	TATCTGCGTT	CTTATCTGCG	AAATGGCCAC	CTCTTTGCTC	5040
TCAGTCTTCG	TCTTCTCTTG	CTTCTCTTGC	TGGCGCAGGT	TTTTATCGAG	CAAGCTTGGA	5100
TTGCGACAGC	AGTGGTAGTT	CTCTTTAACT	ACCTCTTGCT	CTTCCAGTTG	CTGGCCCTCT	5160
ATCATGCCTT	TGACTACCAG	TATTTGACCC	AACTCTTTCC	GCTGGACAAG	GGGCAAAAGG	5220
AAAAAGGCTT	ACAGGAGGTA	GTTGAGGAT	TGACCAAGTT	TGTTTTACTT	GTGGAATTAG	5280
TTGTTGGGTT	GATTACCTTC	CAAGAAAAAC	TAGCCCTTCT	AGCCTTACTA	GGAGCTGGTT	5340

1148

TGGTTTTACT AGTCTTGAT TTGCCTTATC AGGTAAAACG TCAGATGCAG GACTAACATT	5400
GCTGATACGA CACTAAAAAA GAAGTTGAGT TCAGTCTGTC TCAACTTCTT TTTTGTACT	5460
ACAGGATAAT GGTTGGTCCG TAGAGACTTA TACTCTTCGA AAATCTCTTC AAACCACGTC	5520
AGCGTCGTCT TACCGTACTC AAGTACAGCT TGCGGCTAGC TTCCTAGTTT GCTCTTTGAT	5580
TTTCATTGAG TATTAACCTG GTCTTGACTT GGTCAAAGTG GAAGCGGTCA TAGGCCCGCC	5640
AAGCGGCGCG AGTTGGAGCA TCTGGATCAA GAGCGCTGAG TCCCATGAGA AGACTGGAAG	5700
TCTGGTAAAA TTTTCTAGT TCAATCAAGA ATCGATTATC CACTGTTTCA GCCTTGCGTA	5760
GAAAACCAAG AATAGAGTTT AATTGCTCCT GAAAGCGGAC GTCGTCAGCG CTGCGCTGTT	5820
TGCATGCTTG GTAGGCTTTG TTAAAGTCAG TAATCAAAGT ATGAGCTCTT TTGATGGGGT	5880
CTGTATCTGT CATGGGAATG CCTCCTTTAA TCTGGGTGCC AGTCTTACTT CTGGCAACTG	5940
TGTTTTGATA CTGTTAGTTT ATCACTTTTA ATTCTTTTTT TTTATTCAAA TCTTTAATTG	6000
TCATTGAAAT GTCTTGAATT GCGCTGAGTG AATTTTATGA TAAATAGTT GTAAGCTCAT	6060
CATGATGTTG TAGAAAATAA TCCTTTTAGG AGTTTTCAAA GACTGTTTAG GATTGGGTGT	6120
GCTTGGGCTA GACCTTTTCT GTTATTCTTT TCTTAGGAGG AGAATCCAAT GAAATATATG	6180
ATTATTGAGA CGCAGAAAAC AGTCTATAAA GTAAACATCG ACGATATCTA CTATATCCAA	6240
ACACATCCAA CTAAAGCCCA TACCGTACAG ATTGTTACAG AAGAAGCTAG TTTTAATATG	6300
CTTCAAAATT TAAGTAATCT TGAGAACCAA TGTGGGGAAA CCTTGATGAG ATGTCATCGA	6360
AATTGTTTGG TTAATCTTGA TAAATTAAAA TCGATTGATT TTCAAGAAAG AATCCTTTTT	6420
CTCGGAGAAG AAGGTCAATA CGCTGTCAAG TATGCCAGAC GTCGCTATAG AGAAATTCGT	6480
CAAAAATGGT TGAAAGAGGG AGAGTAAGAA GATGAGAATA TTTGTTTTAG AGGATGATTT	6540
TTCCCAACAG ACTAGAATTG AAACGACGAT TGAGAACTT TTGAAAGCAC ATCATATCAT	6600
TCCTAGCTCT TTTGAGGTAT TTGGCAAGCC GGACCAACTG CTGGCTGAAG TGCATGAGAA	6660
GGGGCCCCAT CAGCTATTCT TTTGGATAT TGAGATTCCA AATGAAGAGA TGAAGGGACT	6720
GGAAGTGGCT AGAAAGATTC GGGATCGGGA TCCTTATGCC CTGATTGTCT TTGTGACGAC	6780
TCACTCGGAG TTTATGCCCC TGTCTTTTCG CTACCAAGTG TCTGCTTTGG ACTACATTGA	6840
TAAGGCCTTG TCAGCAGAGG AGTTTGAATC TCGGATCGAG ACAGCCCTCC TCTATGCCAA	6900
TAGTCAAGAT AGTAAAAGTC TGGCGGAAGA TTGCTTTTAC TTAAATCAA AATTGCCCCA	6960
ATTTCAGTAT CCTTTTAAAG AGGTTTACTA TCTCGAAACG TCGCCCAGAG CCCATCGTGT	7020
TATTCTCTAT ACCAAGACAG ACAGGCTGGA ATTTACAGCG AGTTTAGAGG AGGTTTTCOA	7080
GCAGGAGCCC CGTCTCTGCG AGTGCCACCG CTCTTTTCTC ATCAATCCTG CAAATGTGGT	7140

1149

GCATTTGGAT AAGAAAGAAA AACTGCTTTT CTTT

7174

(2) INFORMATION FOR SEQ ID NO: 190:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3207 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

CCACCAGGGA AAATCATTGA AGTTGGTAGT CACCAAGAGT TAATGCAGGC GCAAAGTTTC	60
TACCATCATC TATTCAATAA ATAAGGAGAA TGTCATGAAT CCTAATCTTT TTAGAAGCGT	120
CGAGTTTTAT CAGAGACGTT ACCATAACTA TGCACAGTG TTAATTATAC CTCTTTCATT	180
ACTATTTACT TTCATCTTGA TTTTCTCCCT TGTTGCCACA AAAGAAATTA CTGTTACTTC	240
CCAAGGAGAA ATCGCCCCTA CAGTGTCAAT GCCTCCATTC AGTCAACCAG TGATAATCCT	300
ATCCTAGCTA ATCATTTAGT GGCAAATCAA GTAGTTGAAA AAGGGGACTT ACTCATCAA	360
TACTCTGAAA CAATGGAAGA AAGTCAGAAA ACTGCCTTAG CAACTCAATT ACAAAGACTT	420
GAGAAGCAAA AAGAAGGACT TGGAAATTTG AAACAAAGCT TAGAAAAAGC GACTGATCTT	480
TTTTCTGGCG AGGATGAATT TGGCTACCAT AATACCTTTA TGAATTTTAC TAAACAATCC	540
CATGATATTG AACTGGGTAT CACAAAGACT AACACCGAAG TTTCAAATCA AGCTAATCTT	600
TCCAATAGCA GTTCATCAGC TATTGAACAA GAAATTACAA AAGTTCAACA ACAAATTGGA	660
GAATATCAAG AGTTGAGAGA TGCTATCATA AATAACAGAG CACGCTTACC AACTGGCAAT	720
CCGCACCAGT CAATTTTGAA TCGTTATCTT GTAGCCTCAC AAGGACAAAC ACAAGGAAC	780
GCAGAGGAGC CATTTTTATC TCAAATTAAT CAAAGTATTG CAGGTCTTGA ATCATCTATC	840
GCAAGCCTCA AAATTCAGCA AGCTGGTATC GGAAGTGTAG CAACTTATGA TAACAGTTTA	900
GCAACCAAAA TTGAAGTACT CCGCACTCAG TTTTACAGA CAGCCTCACA GCAACAATA	960
ACTGTGGAGA ATCAATTAAC AGAATTAAAA GTACAACTAG ATCAAGCCAC ACAGCGTTTG	1020
GAAACAATA CCTTAACCTC CCCAAGTAAA GGTATCGTTC ATCTGAACAG CGAATTTGAA	1080
GGTAAAAATA GAATTCCAAC TGGTACAGAA ATTGCTCAA TATTCCCTGT CATCACAGAT	1140
ACAAGAGAAG TACTAATCAC TTAACAGTA TCTTCTGACT ATCTACCTCT ACTAGATAAA	1200
GGACAACTG TAAGATTAAG ACTGGAGAAG ATTGGAAATC ACGGCACCAC CATCATCGGC	1260
CAACTTCAGA CAATTGATCA AACTCCTACC AGAACAGAGC AAGGAAATCT CTTTAAATTA	1320

1150

ACCGCTCTTG CAAAACTATC TAACGAGGAT AGTAAACTCA TCCAATATGG CTTACAAGGT	1380
CGCGTCACTA GTGTAACACT AAAGAAAACA TATTTTGATT ATTTCAAAGA TAAAATTTTA	1440
ACACATTCTG ATTAATTTTC AGATAACACT CTATAACTAT TTATTATCTT ATCAAAAAGG	1500
AGAATCATAA CATGGATAAG AAACAAAACC TAACTTCATT TCAAGAACTA ACAACTACCG	1560
AACTCAATCA AATTACAGGT GGAGGATTGT GGGAAGATTT ATTATATAAC ATTAATAGAT	1620
ATGCTCATTA CATCACATAA GAACTTCATC ATCCAATACA ACTATAAAAA AATAAGACCG	1680
AGAAACAAGT ACTCTCGGTC TTATTTTTC TCACTCTGTA TGTATCACAG TAAGTACCTG	1740
ACGAAAGACT TGATTTTGAC AGGTGGTATT TAGACTGGTA TTAGGATGGC TTTCCACAAT	1800
CTTCATGACG GTATAGAGAC CAACTCCTCT CTCCTCCCTT TTAGAACTGG CTCCAAAGGA	1860
GAAGATTTCA GAAATATCGA TGCCCTCTTC TTTGATGGAG TTTTCGATGA TAAAGGTCTC	1920
CTGTGCTCCA TTTTAAAA AGGCGATTGA AACATGAGGT TGAAGCTT CCACACTGGC	1980
TTCAATAGCA TTGTCACAAA GGATAGACAC AATGGTTAGA AAATCAAGTA GACTCATCCC	2040
CTCGACCTGA ATCTCCTCAG GAACTTCGAC ATTAAAGACA ATGTTCTTAT CTCTGGCTTT	2100
TAAAAATTC CCTGCTAGAA GACTTTTGAG GGCTTTATCA CGAATATTTA CCAATCTGCC	2160
CAGGTCATAT TTATTGTTCT GCAATTTCTG ACTGGAATCC TTAAAGACGG AGCCATAGAC	2220
CTCTTTTATC TGCTCCATAT CCTCCTCTTC AATGCCGAGA CGTAAGCTAG TCAAGAGGTT	2280
GGTATAATCA TGACGAAAGC TCCGTACTTC CTTGTAAAGC TCCTCTATAT GCCGACTATA	2340
GCGTTCCATA TCTCTATAGC GCAGGGCCTG CTCTTGTTCC AATCTCTCAT AGAGTTTTTC	2400
CTTCAAATAG GTATCCAATT TCTTGATAAC CCCCATAAAA AAGAGTAGGT AAAAGACTAG	2460
GATGAGATGG CGAACAGTCT TTGATTGAAT ACTTTGTTCA TATTCAAAAA AAGACAGACT	2520
TTCCATGACT AGATAGTAGC CACCCATTAT CCAGTTAATC TGAGTCAGGG ACTTTTGAAA	2580
GGCTTTATCG AGAATCTCCT TTCTCAAGCT AGTAAATCG TAGTCCAACC ATTTCAAAAA	2640
AGCTAGAGAA ATGAAGAAAT TGAAAATTAT TATACATAAC CCAGTAAATG AGTAGCCATC	2700
ATATACTTGC CCTTGTCCTA AAAATGGAAG CACAAAATAG GAGACTCCTC TATAAAGAG	2760
ATTACCAAT ATCATTGGAA AGAGACCATA AAAGAAAAGG AGTTTTTTAG GAAGCCCTCT	2820
CAATAATAAG AAAGATAAGC CTATGCCGTA CAAGGGTTCC ATAAAATAAG ATAGGTAAAC	2880
ATTTCTACT ATATAGCTAA TCATCACAAA AACAAAGGCC AACAGTATCT TCAAAAGAAA	2940
GGCCTTAAAA ATCCTCTCGA AAGTAAGATC AATTCATCC ACCTTAAAGA AGATGACAAT	3000
TTCTAGTCCA TTAGTAACAA GTGTATACAA CAATATCCAA GCAATGTTCA TAAATTCTCC	3060
TAGCTCAGTG TAATTTATTG ATGGCCTCAG ACACTTCCTT GACCTTATAA CGGGCGATTA	3120

1151

GACAACTTCC ACCATTGGGA GAGAAGAGCA GTTTTCTTT CTTATCCAAA TGCACCACAT 3180  
 TTGCAGGATT GATGAGAAAA GAGCGGT 3207

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10357 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

CTGAATCAAG TGTACTGCAC CAGTTCGTGC ATCAGGCATA ACAACATCTA CAGATATAAT 60  
 ATTGTTTTCT GAGTCCGCCT CATAAGTTAA AATCATAAAT TTTTCGATAT TCGAATTTTT 120  
 AGTAGCTTGT TCAATTCTTT GAATCATTTT ATCAGAAACT AACTCCATCT GAATTGGAAA 180  
 GGAATGACTA TTTTCATCAT TTTTGTAGGA AGAATGTTGA TTAAGATAAA GTGTATTCAT 240  
 CTGAGCATAT TCAAATAAGT AGCCACTCTT ATTTTTTTGT ACCAAAGGAA ATTGGTTTTG 300  
 AAGTCGCTTC TTACCCTTTA TAATTAACAA TACTTTCCCA TATTTTCTG TATTGTTTC 360  
 AAATCTAAA TATCCCAAG TCTGTCCTGC TAATTGTAAT TTATACTCAA ACAAATCTGC 420  
 TGATGCAAAAT GCAGTATCAA TATGATTAGG TCGCGTCCAT GCATAACCAT TCGACACTAT 480  
 CATGTCTCT CTTTTTCTA GACGTTTCATC TACATAATCT TTTTGCCCTT TCATCAAAGT 540  
 ATCTACAATT TTTTGTGCCT CAAGCGAATC AAAGAGATCC TGATTCAACA TAATCTTCC 600  
 TCCTCCAAAT ACTTTTAAAT GAATTATACC ATTTTCTTAA AGAAATTACT ACAATAATTA 660  
 TCTTTTTCTT AAAGTTCTGT GTCAGAGTAA TTTAGAAAAT TATATCTTCT ATAGTAAAAT 720  
 CAATTAAAAA CTGAACAAAT TTATTGGGAA ATTCAAATCG CTTTCTGAAA ATATTTTAGG 780  
 AACCCTAGTG TAATATTCCA GATTCAATTC ACTATAAAAC TGACCTTTCT CTGCAAAAG 840  
 AAAAAGGAAA GACTTCCTTT CGTGCCTTTC CTCTTACTTG CTACTTGTTC GATTATTTTT 900  
 GGTAAGCTAC TGCTTGTCTG ATAAAATCCT GAATCGGCTC TCCTTGGTGG AGAGCTTTTA 960  
 CTATTTTCGA ACCGACGATA ACACCATCTG ACACCGCATT GAAGCGTTCC AGATCGGCTT 1020  
 GACTAGATAC ACCAAAACCT GTCAAGACTG GGATGTCGGC CACTTGATGA AGTTGCGCCA 1080  
 AGTGCTTGTC CAAATCTGCA CGGTAATTGC CTGATTTCCC TGTCACTCCA TTGATGGCAA 1140  
 CGGCATAGAT GAATCCCTCC GCCCCTTCAA TCAACTCTTT CTGGCGCTCA ATTCCTGTGG 1200  
 TCAAGCTTAC TAAAGGAATC AAGGCGATAT CTGTATTTGC CAAAAATGGT TCTACAAAGT 1260

1152

TGGCATGTTT	ATGAGGCAGG	TCTGGGATAA	TCAAGCCCTT	CACAGCTGTA	TCAGCCAGAT	1320
CTTTGACAAA	GTTCTCCACA	CCGTACTGAA	AGAGGGGGTT	GAAGTAGGTC	ATGATGACCA	1380
GTGGAATCTC	TGTTTCAATG	GTTTTCAAGG	TTTCAACTAA	AGCCTGGGTA	GAGGTCCCGT	1440
GGGCTAAACT	GCGCAAGCCA	GCTTCTTCGA	TAACAGGTCC	ATCTGCAACA	GGGTCTGAAA	1500
AGGGAATACC	CACCTCAATT	GCAGAGACAC	CCAAATCTTC	TAAAAAGTGA	ATTGTTTCAG	1560
CAAGACCGTC	CAAACCTTTC	TCGTGGTCAC	CAGCCATGAT	ATAGGGAACA	AAAATTCCTT	1620
TTCCAGCTGC	TTTAATAGCA	TTTAATTTT	CTGTTAGTGT	CTTAGGCATG	AGCTTCTCCC	1680
TTCTTTGCTG	CATCTGCTTC	CAAGCGGTCC	TTGACTTGAA	CCACATCCTT	GTCCCCACGA	1740
CCTGATAGGC	AGACAATCAT	AGACTTTTCT	GGTCCAAGTT	CTTTGGCCAA	TTTCACCGCA	1800
AAGGCGATAG	CATGGCTAGA	TTCCAAGGCT	GGGATAATCC	CTTCCACACG	AGACAAGAGT	1860
TGGAATCCTT	CCAAGGCTTC	TTCTGCTGTC	ACAGGGACAT	AGCTGGCACG	TTTAATATCG	1920
TGGTAGTGAG	AATGCTCTGG	ACCGATACCA	GGATAGTCCA	AACCTGCTGA	GATAGAGAAG	1980
GCTTCAAGAA	TTTGACCATG	GGCATCTTGG	AGCACATCCA	TGAGGGAACC	GTGAAGGACA	2040
CCTGGACGAC	CCTTGGTCAA	GGTAGCTGCG	TGGTGCTCTG	TATCCACACC	AAGCCCTGCT	2100
GCTTCAGTTC	CATACATAGC	TACTGACTCA	TCTTCTACAA	AGGGATGGAA	GAGCCCCGATA	2160
GCATTGACCC	CACCACCAAC	ACAGGGTACT	AGGGCATCTG	GCAGATCTCG	ACCTGTCAAG	2220
TCACGGTACT	GTTGTTTAGC	CTCTCGACCG	ATGACACTTT	GGAAGTCACG	AACGATTTCT	2280
GGAAATGGAT	GAGGCCCCAA	GGCAGAACCA	AGGATATAGT	GGGTATCGTC	GATATTAGCC	2340
ACCCATGAAC	GAAGGGCTGC	ATTGACCGCA	TCCTTGAGCA	CGCGCGAACC	ATCTGTTACA	2400
GCCTCGACCT	TGGCTCCCAA	AAGCTCCATG	CGGAAGACAT	TGAGGGCTTG	GCGTTTGACA	2460
TCTTCTTCAC	CCATGTAGAT	GGTACATTCC	ATGTTAAAGA	GGGCTGCAGC	AGTTGCAGTT	2520
GCCACACCGT	GCTGACCAGC	ACCCGTTTCT	GCGATAATTT	TCTTTTACC	CATGCGTTTG	2580
GCAAGCCAAA	CTTGTCCTAA	GGCATTGTTA	ATCTTGTTGG	CTCCTGTATG	GTTAAGGTCT	2640
TCCCGTTTGA	GATAAATCTT	GGCTCCGCCA	ATATGCTGGG	TCAAGTTTTT	TGCGTAATAA	2700
AGAGGAGTTT	CACGTCCTAC	GTAAGTGGCG	AAAAGCTGGT	TTAATTCCTC	TTGGAAACTT	2760
GGGTCTGCCT	GACTTTCACG	GTAGGCCTTC	TCCAACCTCA	AACTGCTGT	CATCAATGTT	2820
TCTGGGACAA	AACGTCGGCC	GAATTTCCG	TAAAATCCAT	CTTTATTTGG	TTCCTGATAT	2880
GCCATGCTTT	ACCCTCTCTA	TAAATCTTCT	AATCTTTTCA	TGATCTTTTT	GTCCATCTGT	2940
CTCCACTCCG	CTCGATACAT	CTACTGCATA	GGGAGTAAAG	TGTTGAATTG	CTTTTACTAC	3000
ATTATCTTCA	TTAAGGCCAC	CTGCGATAAA	GAAGGGCTGT	GCTAGTCCAG	TCGTATCCAG	3060



1153

TTGACCCCAA TCAAAGGGCT GGCCACTTCC TGCCACAGGG GCATCAAAGA GTAGATAATC	3120
TGCCTGAGAA TTGGGGACAT GCCCATTTC ATCTACCTGC ACAGCCTGAA TACTGGCACA	3180
AGGCAAATTC TCAAATAAAT CATCTGCCAC CTGACCGTGA ACTTGAACCA AGTCCAAGCC	3240
AACTTTGTCA ATCGCTTCCA GCAGTTCTAC CCGACTTGGT GAAACAAATA CTCCAACCTT	3300
TTTCACATCT GCAGGAATAA GCTTTGCCAA CTCAGCTGCC TCTTCTAAAG TCACCTGTCT	3360
TTTACTAGGT GCAAAGACAA AACCGATATA GTCGGCTCCT GCTGAAACGG CTGTTTCCAC	3420
CGCTTCTTTG GTCGATAGTC CACAAATTTT AACCTTTGTC AATCTGCAAC TCCTTGATTG	3480
TCTGGGCCAC ATTTTCTGCC TGCATAAGAG CTGTCCCTAC CAAATTCGG TTAAGTATG	3540
GGGCTAGTCG TTCCGCATCC TGCCCTGTGA AAATGGCAGA TTCAGAAATG TAATAGCGAC	3600
CTTCTCAAA GTAAGGGGCT AAATCTACAC TGGTCTGCAA GTCGACCTCA AAGGTAGTCA	3660
AGTTGCGGTT GTTGACCCCG ATAATCTCAG CACCAAGTCT GTGGGCTACC TCTAGTTCAG	3720
CTAGATTGTG AGTCTCCACT AAGACTTCCA GACCAAGCTC TGTCGCGTAG TCATACAGTT	3780
CCTTGAGGCG TTCTTCGGAC AAGGCTGCCA CAATGAGCAA GATAACTGTC GCACCTGCAT	3840
TGCGAGCGCG GATGATTGTC TTTTCATCGA TGATAAAGTC TTTGTTGAGC GTCGGAATCT	3900
CTACCTGACT GGAAATTTCC CGTAGATAAT CCAATGCCC TTAAAGAAA ACCTCATCTG	3960
TCAACACCGA AATCATCACT GCTCCGTTTT CTTCATAAGT CTGGGCCTGT TGCACAATAT	4020
CCACATCGAG ATTGATATCT CCCAACTAG GGCTAGCTTT CTTGACCTCA GCGATTACCT	4080
GCAAGCGGTC CTGATGATTC TTCAAAAT CTGCCAAGCG ATAGGTCTGG CGCAGAGGCT	4140
GGATTTGCTC CAGCTTCATC TGCTCCACCT CACGCGCCTT CTGCTCTAAG ATTCGTGCTA	4200
AAAATTCCTG ACTCATTTTT GGTACTCCTG TAACAGTCTG AGTTTTTCAA GGGCCTTGCC	4260
TCTAGCAATC ACTTGACGGG CCAAGGCAAC CCCTTCCTTG ATGCTATCAA TCTTACCATT	4320
AGCATAGAAA CCAAGACCAG CATTCAAGAC TGTCGTTTCC AAGAATGGAC TTGCTTCGTT	4380
TTTCAGAACG CTAAGCAAAA TTTCTGCATT TTCTGAGCA TTCCCACCAC GAATATCTTC	4440
CATAGCATAG CCTTCCATTC CCAAATCCTC TGGAGTAAAG CTTGACAAGC TGATTTGCCC	4500
ATTTTCAAGA AGTGCAATCT TGGTTGTTCC GTTCAAGCCA GCTTCATCCA ACCCTTCTGG	4560
TCCAGCAACC ACGATGGCAC GTTTGCGACC CATATTTTTC AAAACCTGAG CTGTACTTTC	4620
TAGGAGTTCT GGACGACTAA TTCCAAGAAG CTGTGTTTCT AAAGCCATTG GATGAATCAG	4680
TGGACCAGTC AAGTTCATAA TCGTTGGAAT TCCCAATTCC AAACGAGCTG GCATGATGTA	4740
TTTCATAGCT GGGTGCATAT TTTTAGCGAA GAGAAAGACG ATTCCAGTTT TATCAAAGAC	4800

1154

CTTACCTAGT TCAGCTGGTT TGAGGTCAAG ATTGATTCCC AAGGCTTCGA GGACATCTGC	4860
GGAACCAGAT TTAGAAGATA TCGAGCGGTT ACCGTGTTTG GCCATGTGAA TACCGCCACC	4920
AGCCAAGACA AAGGCTGCAG TTGTGGAAAT ATTAAACTG AAAGACTTGT CCCACCTGT	4980
ACCACAGTTG TCCATGGCAT CATGAATCTC AGTTGGAATA TGCTGGGCAT GTCCTCTCAT	5040
GACTTGGGCA ATGGCTGTGC GTTCTTCAGG TGTTCCCCC TTCATCTTAA GAGCTAAGAG	5100
GAGAGAAGCA ATCTGCGCTT CAGTTACACG CCCAGTTACG ATACGCTCAA TGACATCCGT	5160
CATTTCCACA CCTGATAAAT TTTCAAATTT TGCTAGTTTT TCAATAATCT CTTTCATCCT	5220
AGTTTCCTCA CTTTACAACC TCCTCGATAA AATCCGAAT AGAAGACAAG CCGTCTGGCG	5280
TTCCAATGCT CTCTGGATGG TACTGGAAGC CATAAATCGG TAGGTTTTTA TGTGAATCC	5340
CCATGATGGC TTGGTCATCA GTCGAACGAG CTGTCACTTC AAAGTCTTCT GGCATTTCTT	5400
CAATCAAAAT ACTGTGATAA CGCATGACCG CACGGCCATC CTCAATACCT TGATACAAAA	5460
CAGATGGCGC TTCAAAGTTG ATATTGCTCT GTTTCCTCATG CATGACTTTT GGAGCCAAAC	5520
CTAGCTTACC ACCAAAGACT TCTGCAATGG CTTGGTGGCC CAAACAAATC CCAAGAATCG	5580
GCTTCTTGCC TGCAAAATCA CGAATCATGT CTTCCATCTT TCCAGCATCA ACTGGCCAAC	5640
CAGGACCAGG AGAAAAGACC AGACCATCTG CTTTTTCAGC TTCTTCATAC AGCTTGAAT	5700
CATCATTTCT CAGAACCTGA ACTTCTGCAA AATCCCAAT GTATTGGGCC AAGTTATAGG	5760
TAAAAGAATC ATAGTTGTCA ATCAATAAAA TCATGGTCTT AGTTCTCCAA TTCTAGTCAT	5820
AGATTTTGCT TTGTTAATGG TTTCTTGTA TTCGTTTTGG GCGATAGAGT CGTAGACAAT	5880
CCCTGCCCCA GCCTGCACAT AGGCTCTTTG ATTTTGTAGA ATCATGGTTC GGATGGCGAT	5940
GGCCAAATCC ATATCACCCG TCGCAGACAA GTAGCCGATT GCCCCAGCGT ATACTCCCCG	6000
TTTTTCCGTT TCCAGTTCAT AGATACGTCT CATCGCTCGA ATCTTTGGTG CTCCAGAAAC	6060
GGTTCAGCA GGAAGCGTTG CTTTCAAGGC ATCCATGGCA GTGAGTTCTG GAAGCAAACG	6120
CCCCTTGACT ACGCTGGTCA AATGCATGAC GTAGCGGAAG AGCTCCACTT CCATATACTT	6180
AGTGACTTGG AACTGGTTCG TTTCAAGAT GCGGCCAATA TCGTTACGCC CCAAGTCTAC	6240
CAACATTGCA TGTCTGCTG TTTCTTCTC ATCAGAGAGG AGGTCAGTCG CCAAGGCCTT	6300
GTCTTCTTCA TCCGTAGCCC CTCTGGTTCG CGTCCCTGCA ATCGGATTGG TTGTCACGAT	6360
GCCATTTTTG ACAGAAACCA AACTTTCTGG ACTAGCTCCG ATGATTTGAT AATCCCCAAA	6420
ATCATAGAAA TAAAGGTAAT TAGAAGGATT AGTCACGCG AGATTTCTGT AGAAGTCAAA	6480
TGGATTTCCA GTAACCTCTG CTGAAAAACG CTGGCTGAGT ACACATTGGA ACATATCTCC	6540
GTTACGAATC AAGTCACGAG CTGTTTCTAC CATTCCTCA AACTTATGTG GAGCGATATG	6600

1155

CGGTTTGAAG TCTAACGGAG ATAGATCCAA ATCTTCAAAT TCATTGGAG CAGGAATGCG	6660
TAATTCCTCA AGCACTTGGT TCAAGGATTT TTCCAAGGCC TCTTGACTGC GCTCACTATA	6720
AAGTGCATCC TCTATGACAT GTATCTTCTC CTTCTTGTGG TCAAAGACCA TATAGCTCTC	6780
ATAGACAAAG AAATGCATGT CTGGCGTCCC AATTGTATCC TCAGGGATTT GACCAATTTT	6840
TTCATAAAGC GAAATCATAT CGTAACCCAC AAAACCAATG GCTCCACCAC CAAAAGGTAG	6900
CTCTGAGTGG TGCTGACTCT TATGAATCAC TTCATAAAGG AAATCCAAGG GATCCCGATC	6960
AATCACTTGA CCATTTTGAT AGAGAACCCC ATTTTCAAAC TTAATCTCAA AAATGGATT	7020
ATAGGCTAGG ATAGAAAAAC GAGCTGTTTC CTTGTCTCTC GGAATACTCT CTAAATAAC	7080
CTTATGTTGC CCCTTTAAGC GCATATAAGC CAAGATTGGT GATAAGACAT CTCCATGAAT	7140
GATTGCTTCC ATTGTAATTT CCCTTTCAGT TCTACTTCTA GTCCGTGGTG ACTGTATGAA	7200
AAATCCCCAC GCAAAATAAC TTGCGTGAGG ACGAAATTCG CGGTGCCACC TCAATTATAG	7260
GATTTCTCCT ATCTCTCATT CCTGTCTCAG ATATCTCCTG TAACAGGCTG TGCGATAAAG	7320
GGCACTCCCT TGAGAATGAT GTTTTCTTCT CTCGTTTCAG ATGAACCCAA CTTTACAGCT	7380
TTCTCTGCTT GTTTTCAGCA ACCACAAGCT CTCTGTGAGA GAAAGAACTG TAATTTTTCC	7440
ATCTATTATT TTTTAGCTTC TAGTAGTCTG CAATCGCAGC TAGGTCCTTG CCTCCACGAC	7500
CAGAGACATT GATGAAGAGA TGTTCATCTC GGTACACCTT TATACTCTTC GAAAATCTCT	7560
TCAAACCGCG TCAACGTCGC CTTGCCGTAG GTATGGTTAC TGACTTCGTC AGTTCTATCT	7620
GCAACCTCAA AACAGTGTTT TGAGCTGACT TCGTCAGTTC TATCCACAAC CTCAAAACAG	7680
TGTTTTGAGC TGACTTCGTC AGTTCTATCC ACAACCTCAA AACAGTGTTT TGAGCTGACT	7740
TCGTCAGTTC TATCCACAAC CTCAAAACAG TGTTTTGAGC AGCCTGCGGC TAGTTTCCTA	7800
GTTTGCTCTT TGATTTTCAT TGAGTATTAC TAGCTTTTTT CGTATTAGTC CAGCCTTTTT	7860
GTTTGCTTTT AGTAGTAGGC ATGGAGCTGT AGATAGAACT CAAGTTCATC AAAGCGACTT	7920
AAGGCCCTAA TAAAAGATAA ACCAAACGAC GGATAGAAAA AAGCCACAC ACAGAATATA	7980
CTCCCGTGTG AGGGCGTTGG TAACGCGGTG CCACCTCAAT TATAAAGGGA CTATCCCTTT	8040
ACATCTCTGC CTTGTTTAAC AACAAGCTGC ACTGTAAGGT GTGCGCACCG AATTTTCATT	8100
GTTTCAAATT CATTTTCAA ATCAGCCAC TTCTACTACT TCCAACCACC TATTCACAAT	8160
CACCACAGGC TCCCTGAAGA TCAAAAATAG TTACTTTTCT GATTTGTTGA ACTTATTTTA	8220
ATACTTTGTT TTTCTTTGT CAAGACTTTT TTACGATTTT TTTGAAAATA TCATTGGAAT	8280
ATGACCATGT CTCCTTAGA TCGAACATGA ACATGTCCCA CTTCTTAGAA ATTGGATCCA	8340

1156

ACTCAATAGA AACTGAATGG AGGCTAAACA GAACTTATTT TAGAACACTC CATCTTTTCC	8400
ACTAGGATTT TCAAGAATTA AACAACTACTA GAACTCTGT CTCCTAACAA ATTTAGGAGA	8460
AACTTCAACA GATGTGACAC TTTCCCCTTT AATAATTGCT AAAACACCTT CTATCATTTT	8520
TTTAGCCAAT TTAACATAAT TGGGAGCAAT TGTAGACAAA GCTGGAGTAT AATACTGAGA	8580
AATAGGAATA TTATCAAATC CAATGATAGA AATATCATCT GGAATAAGAA TTCCTTTTCTC	8640
ATAGCACGCA CGAATCAAGC CCTGAACCTT TTCATCTCCT GAAACAAAAA TAATGTCCGG	8700
ATAATTTTGG GTAGTCAAGT GCTGCATTGC ATAAGAATAA ACTGAATCAA TTGTAGATAA	8760
GCCATAAATG ACTTTTAAAT CCATAAAGTA ATTTTATCA TTCAGAAAAG AACGCACACC	8820
TCTTTCACGA TCCTTATTAA CATGGGATTC TCCTCCCATTA AGCAACCACA TATTTTAA	8880
TTTTTCTTCA GTTACAGCTT TCATCATATC ATAAGTAGCT TGAAAATTAT TATTAGATAC	8940
ATAGACTACT CCAGACGTTT GAGATTACCC GAAAACAAGA AAAGGCATAT GGTCTTCTT	9000
TAAATACTGA ATTCTGATAT CATCTACACT TTCATAAAAA ACAATAACAC CATCTACTAG	9060
GCTACCTGTG CTTGATATAA TTGAATTACT AATTGTATCC TCCTCTCCAA AGTACTCAAC	9120
TATAGCATTG ACACCAAATT CTTTACACGT CCGTAACACT TTATCTAACA GCGTATGAAA	9180
CCAAATTAAG GGAAGAGT CGATTTTTTT TACAGAAATC AATATATTTA TAGCTTCTTT	9240
TTTAGTTAAA TTTTTGTCAT ACGCATTGGG AATATACGAC AATTCCTCTA TAACTTTTGT	9300
AATCGCTTGA TAAGTTTCTT CTTTAACATT TACTCCACCA TTAATAACTC GTGAAACTGT	9360
TTTTGGAGAA AAACCTGATA AACGTGCAAT ATCATAAATA GTTACCTTTT TCCCATTAT	9420
ATTTTTCATT TCAGTCTCC ATTACGAACA TTCTAATATT ACTATACAAT ATTTAATTTT	9480
TTTTAACAAG AGAATTTAGT AAATTATTTA AGATCCACAA ATTCACAAAA TTAATTTTAC	9540
AAATATCTT CCCCTTCAAA AAAGTTTAAA TTGCATTCA CACCTTTATT TTTAAGAATG	9600
TTTCCAACCT CACGACAAAT AAATTCATAT GAGAAAAAAC TGCCATAAAA TTGTAGATTA	9660
ACTTTTTCAG TAAAATGTGT AGGATTTATA AAAACATATA ATAGCCTGTC AATGTAACAT	9720
TTTAACATAG AGTTAATTTT TTCTTTAAAG ATAACATTTG TTATCAACTC ATCAGGAGGT	9780
AAATGAAAGG CAAACACCAT TTCACAAATA TCATAAAAAG AAATAAATTT GTATACTTGT	9840
ATCAACAAT TATTATCAAA ATATTCTATT TTACCTAAAT CAAAATTGAT TTTATAATCT	9900
TTCATAAAAA CCTCTGAGCA AAAATCTACT CAAAAATTAG ATGATTAAAA CATCTAAAAA	9960
GCAAAAGGAC AAAAACATCT GTCCCTTTGT TTAATAAATT TCAGCTAATT TCTTCGACAT	10020
AAATAACACC TACAATATTA GCAATTTCTT CCATCAGTCG AAGATGTTCA AATCTACCTG	10080
ATAATTCCAG AGTAATAAAT GACGCTATTT TTTGTCCGG AACATCAAAG TATTCAATTC	10140

1157

TGTCAGAATT AACATCTCCA AACGCTGTTT TTGAATCGGT CATTCTGATA CCATTTTCTG	10200
CACAATAAAC CAATACACGA TTATAGGCTT CTGTAGATTT AACCACTATA TACAATTCAA	10260
TCATTTTAGA ACGATTTTGC AGATATTTTT TTAGTGTTG GAACATGGAT ATCACACCCC	10320
AAACAGAAAT GGCTACTAAA AGAGCTCCCT CATAAGG	10357

(2) INFORMATION FOR SEQ ID NO: 192:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6867 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

CGGGACATTC TCAATCTTCT GTCTTTTGT TTTCTCTTCT TTCTATGATA CAATGGAAAA	60
AATAAATTCA AAAGGAGTTT TTTTATGACT TATCCAAATC TCTTGACCG CTTCTTAACC	120
TATGTTAAGG TCAACACGCG CTCTGATGAA CACTCTACTA CTACTCCAAG TACACAGAGT	180
CAGGTTGACT TCGCAACAAA TGTCTAATT CCTGAAATGA AACGTGTTGG ACTGCAAAAT	240
GTTTACTATC TACCGAATGG TTTTGCTATT GGAACCTTGC CAGCCAACGA TCCGTCCTTA	300
ACACGTAAGA TTGGTTTAT ATCGCACATG GATACTGCTG ATTTTAATGC TGAAGGAGTC	360
AATCCACAGG TAATTGAAAA CTACGATGGT GGTGTGATTG AACTAGGGAA TTCTGGTTTC	420
AAACTCGATC CAGCTGACTT CAAGAGTCTT GAAAAATATC CAGGACAAAC GCTCATCACA	480
ACAGATGGAA CAACCTTGCT AGGTGCTGAT GACAAGTCAG GAATGCTGA AATTATGACA	540
GCCATTGAAT ATCTAACTGC TCATCCTGAA ATTAAGCACT GTGAGATTCG TGTTGGTTTT	600
GGTCCAGATG AAGAAATCGG TGTGGTGCC AATAAATTG ATGCAGAAGA TTTTGATGTG	660
GATTTTGCCT AACTGTTGA TGGTGGTCCA CTAGGTGAAC TTCAGTACGA GACTTTCTCA	720
GCCGCTGGTG CTGAATTGCA TTTCCAAGGT CGTAATGTCC ACCCTGGTAC TGCCAAAGGG	780
CAGATGGTCA ATGCCCTTCA GCTAGCAATT GATTTTCATA ATCAACTTCC AGAAAATGAC	840
CGACCTGAGT TAACTGAAGG TTACCAAGGT TTTTACCATC TAATGGATGT GACAGGTAGT	900
GTTGAGGAGG CGCGTGCAAG CTACATCATT CGTGATTTTG AAAAAGATGC CTTTGAAGCG	960
CGTAAAGCAT CCATGCAATC TATCGCTGAT AAGATGAATG AAGAACTTGG GAGCGACCGT	1020
GTCACCTCA ACTTGACAGA CCAGTACTAC AATATGAAAG AAGTCATTGA AAAAGATATG	1080
ACTCCAATTA CCATTGCTAA AGCCGTTATG GAAGATCTAG GTATCACGCC TATTATCGAA	1140

1158

CCAATCCGGG GTGGAACAGA CGGCTCTAAG ATTTCTTTTA TGGGAATCCC AACTCCGAAT	1200
ATCTTTGCAG GTGGCGAAAA TATGCACGGA CGTTTTGAAT ACGTTAGCCT TCAGACTATG	1260
GAACGTGCAG TTGATACCAT CATTGGCATC GTAGCTTATA AAGGCTAAAA AGACGAGGTA	1320
GCTCAGCTAC TTCGCCTTTC TTTTATTCT ACTGGTTTTT CTGATTTCC AGTAGTTGTA	1380
GAAGATTCTG TTGTTTCATT TTCTGAAGTT GATTGAGCAG GTTTAGAATC TCTGTATTG	1440
CTTGGTTTGT TTTCGTCGCT AGCAGTTTCA ATGTTAGATT CTGCAGTTGC GTTTGGTTGG	1500
TTCTCAGCAC TGGTGTATC ACCATTTGCT TCAGCATTTT TTGCTGGACT TGTTCCTCA	1560
CTTGGCCTAG CTTTGTACTG GATTTGATGA TTCAAACTA GAATAGCTTT TGTGATTC	1620
AGTAAAGCTG TTTTGTCTTT ACTCTTAGCA GAAAGTTGAT CTAATAATGC ATCCACCTTA	1680
TCAAAGTCCG CATCAGATCC ATTATTACTT TCTAAATAAG AGTGAAGCGA CATGAGAATA	1740
TCGTAGAGTT TTTGATAGAG TACAAGTGTC TGAGGATCTT GCTCAGCATT TTCCTTTCT	1800
TGTTGAAGGG CGCTAGCGAT ACGAGTCAAG ACATCTTTTA CCTGACTGTT TACTTCATCC	1860
AAGTCTGCAT CAGCCTTGTT TGTGGCAGCT TTTAGATTTT CTACTTCTTC TGCCAAGGAT	1920
TGTCTGATTC CTTCTTCATG GATTTGTTC AAGAGTTGAT TTGCCTTGCT CAAAAGACTT	1980
TCTACTTCTT CCTTGCTATC TGTCGCAGAT TATTGGTTGC TATCTACCAT GTACTCTAA	2040
AACAGGAGAG TTATAATCCA AGATTACAAG GCCTTACAGA AATAAGAAAT CCAGATAAGA	2100
CAATGTTTCT CCAAGACGCT ATTCGCTTCG CACAGCAGCA CGGATTCAAT ATGCTTTAAT	2160
TTTAAAGTTT AGGTGTCAAG ACCTCTTTTT AGTGTGCCCA AAATTTAGAG AAGTAATCAA	2220
TCAACTAACT TTTATTTTTT TCAAACCTTC AGTAACTGA CCTAAAGCTA ACTCAATCTG	2280
TCTTTGTAGA TGCTTCTGCT ATCAGCTAGA AGTTGATCTA CTTTGGCCAA GACTGCCTTC	2340
TCATCAAAAG TTCCAGGTTG ATAGTTGGAT TGCAGGGATG GAATCTTGTT TTTCAAAGCC	2400
GCTTCATATC CCTTAGTTTG AACCTTGATG TAGTGATTGT GGTGCCCATG AGGAATCACA	2460
AAACCTTCTG AATCTTCACT TATAATTCGA TTGGCATCAA AACCATGACC ATCTTCTTCC	2520
TCATGATGGA CATGTAGTGA CGGATTACTT AATACAGAAC TAGAAGAACT TCCTACCTCT	2580
TCCGTGTTAG AGTGTGATGG GGGATTGTTA AGAGATGACT TAGGAATATA GTGATAGTGA	2640
TCCCCATGTC TTAATATATA AGCATCACCT GTATCTCTGA CAATATCATT AGGGTTAAAG	2700
ACATATGTGG CTGCTAATTC ACCTGCCGAC AAGTCACTCT CAGGAATGAA ATGATAGTGA	2760
CCACCATGTG GTACTATAGT AGATTGAAAT AGAATATGAG CAAATTGATA AGGGGATTTT	2820
AAAGTAATTT CTAACAATGA TTTAGAACT ATGATGTGCT ATTCTAAATT CAACTCACTA	2880
TATATAACCA TCATCGGTAG TATAACGTCC CTGTAATTTT GCTACAGATA CTTCTGCACT	2940

1159

AGCTCCTTTA TCGTCTTTAC CATGTTCTTG TTTTGGCGA TTGATTTCAT CTTTGTTCG	3000
TACATTTTCT GCATGAGCTT GATCTTTAAG GTAAACATAA TACTTTCCAT CTACCTTAAT	3060
AATATATCCT CCCTTAACCT AACTGACGAT ATCTTGATCT TTCGGCTGAT AGTTGGGGGC	3120
TTTCATTAAT AGCTCTTCAC TAAAGAGCGC ATCAAAAGGA ACTTTACCAT TATAGTAGTG	3180
ATAATGATCG CCATGAGAAG TTACATAACC TTGATCTGTA ATCTTAATAA CAATTTGTTT	3240
TGCTTGAAAT CTTCTTTTTT GACTAACCTA GTCTGGAGTC AAATTTTCAG TCTTCTTAGT	3300
GTCTTTATTA CTGTTTACAT ATGAAACACG ATTTTATCT GTATTGGCCT GTTAGCTATG	3360
TTGGTTCAGA GCATAAACAC ACAGACTTAA GGAAAGGATA ACAACAGATC CAGCTGCTAT	3420
ATATTTCTTT TTAATTTCA TAATTACCTC ATTTCTATAA TTATTTATAT GATGCTTCA	3480
TTATTAAATG ATTAAATAA TTAATTAACC AATTAATTAA CTAGTAAATA TTCCACCTCT	3540
TTTAAAGTTG TATGTCAAGA AATTTTATAT ATTAATAATA AAATGAAATT CTCCCAAAGT	3600
CAGAGTTTAA TTTCTAACTT TTGAGAGAAC TTCATTTTTG ATTCAGACTT TTTCTACTGC	3660
TATTCCTTAC GCTATGAGAT CAGATAAATT CTTTTTATC ACTTCTCCAC TTGGCAATCT	3720
TAATTCAATC GTTCCATCCA TATTGAATAT AACACTATCT AAGCCTAATC CGTAACTAGC	3780
TGTAAATTTT TCTAATTTTT CTGTACAGG ATCTACTGCT GGAGCTTCCT CTAATGCTGG	3840
ATCTAACATA GGGTCACTCC CCACATCCC TTCTGGATTC AACATTCCAT TATCCGTTGA	3900
GTTTTCTGGT TTTACAGGTT TTTCGTTGG TGCCTCTGGT AAAGAATCTG CTGGTTTATT	3960
TTCTGTGGT TGGTTCTCAA CTGTTCCAGT AGATACTTTT CCATTTTCAG ATGGTTTATT	4020
TTCAACATTT CCTTGAGGTG CTTCTCCTGT AAAATCTGCC ATATTCTTTT TAATGACTTC	4080
TCCCGATGGT AAATATAATT CAATTGTTCC GTCCATATTA AACAAGACAT TTTCTAGCTT	4140
CATCCCATAA CTTTCAGCAA ATTTTGCTAC TTTTCTTGT ACAGGATCCA CTGTAGGAAC	4200
TTCTTCTAAC GTTGAATTAC TAGTACTATT CCCAGTTTCA GAAAGTTTTT CTTTTCTAC	4260
CTTCTCACTA GTCTTTGGTT CTTCTACCTT TTCATCAAGT TTTAAGTTTT CTTGTGCTTT	4320
ATTCCTTTA AATTGTGGTA GAATACTGG TTTATCAGT TGATTTTCTT TTTCCAAGAT	4380
AGTACTTCC ACAATATAAG TCGATTGATT GTCCAAATAA GCATTTGCCA TGAAGGTAC	4440
AGGAATTTA TTTCCGGCCG TTCTGGTTGT TCCTGGTTT AATTTCGGAA TCGGTAATTT	4500
GATTTACCA ACTTTATAGT TATTTCTAA ATAAGCATTT CCATGAAATT CATCAAACAC	4560
TCTGACTAAA GCATCAGTTC CTTTAGGCAC TGCAAATTGA GGGTCACTC TTAAATAAGT	4620
ATCCCCTGCA TGGAAAGGAT AGAAAATCGT TTGACTGGCC ATTTTGTAAG CTAAAGAGGT	4680

1160

TGGAACGTGA AATGTACCAT CATAACTTAC TTCTGGATAA TCTTTTGAAG CGATAGTATA	4740
CTTAAATGTT TGTCTGGTA AATAAGGTG ATCTAATTCA AAGTTTGCAA TATTCCTAC	4800
TCCTTCTCCA AATACTTTAC CAGATACTTT CTCCAATACT TTCCATCTG GTGTTATTAA	4860
TTTTACTAGC ATATTGATAC CTAATTTTTT CTCCAATTCA GCGGAAAAAC TAAAAGAAAC	4920
GCGTTTTTGA CCATTGGCTA GAGTAAAGTT TTGATTATTA AACGTACTAT TTTTAAACAA	4980
ATTAACAACA TTCGTTAATT CTTCTCCAGT ATAAACTTTA TTCCCTTCTT TTTTAGCAAC	5040
TCCTTCTTCG GGTTTAAACA GTTCATAGTT ACTGTGAGAA TGACCAATTC CAACCGGTTT	5100
ATGTTTCATCA ATCGGATCTG CATGATGGTG ATCTCCATGC GGATAAATAA TCGCATTTTT	5160
TTCTTTATTC ACGACAATAC TTTCACGTTT GACACCATAT TGTTTCATAA TGCCAGCAAT	5220
TTTTTCTTCG ATTTTTTTAT CTAATCTTTT CATTTCTTTG GCATTACTTG GATAATCCTG	5280
TTCATGAGAT GACAAAGAAT CTAATCCATT ATGACTAGTT TTAACCTCCT CTAATGTTT	5340
TTGCGCASCT TAATTTGCTC TTCTGTCAAG TCCTTCTTGA AGAAATAATG ATTGTGGTCT	5400
CCGTGACTCA TGACAAAACC TGATTCATCT TCAGCGATAA TACGATTAGC ATCAAATCCG	5460
TATCCATCTT CTTCATGTTT CTCATGTGAA GTTCCTGGAT TGATTGGAAG AGATGGAGAA	5520
GGTGTGCTA GACTATTGTT TGAAGAGTC GGTGGCCAA TTTGATTGA TTTTGAATG	5580
TAATGGAAAT GATCACCATG TCTTACAATA TAAGCTGTAG CCGTTTCTTC AACGATATCT	5640
TTTGGATTAA AAATATAACC ATCAGATGCT GAAGAGAGCT CCTTACTTGT CGTTAAAGAA	5700
GAAGGATTGC TTGAAAGACT GCCTAGACTA GACACTACTT CATTAGGTTT TGCATTTGTA	5760
GAACTGTAG AACCAGTTCC ACTGATAGGC ACCATTCTGG CAATCTTTTC TTCTAAGGCA	5820
GAAAGCTTGC TGTAAGGAAT AAAGTGGTAA TGGTCGCCAT GCGGAATCGC AACTCCATTT	5880
GGTGTACGAC TGATAATCTT AGCAGGGTCA AAGACCAGGC CATCTGATTC ACTGTAACGT	5940
TGGCGCTAG GTGAATCATA GAGTTCCTTC AAAAGACTCT GGAGATTTTC AGATTTATTT	6000
GCTGGCTTGC TAGTTGATCC TTTTGCTACA GATTGCGTGT TATTGTCACT AGCTGTTGAA	6060
GAATAGCTTA ACTGACTCGG TTGCATATTT TTTCAGCCA GATGTGCTTT AGCTGCTGCT	6120
AATTCAGTAG CAGATAAATC GCTTTTGGGA ATGTAGTGAT AGTGACCTCC ATGAGGAACG	6180
ATATAAGCAT TACCCGTATC TTCGATAATA TCAGTGGAT TAAAGACATA ACCATCATTT	6240
GTCGTATATC GTCCCTGAGA CCTTGCTACA GCAACATTAG AGTTAACCTT CTCATTATCT	6300
TTGACATGTT CTTGTTTTTG ACGATTGATT TCATCTTTAG TTCGAACATT ATCAGCATGA	6360
GCTGCATCTT TCAGGTAGAC ATAATATTTT CCATCGACCT TGATGATATA ACCACCCTTG	6420
ACTTCATTGA CAATATCAGC GTCTTTAAGT TGATAGTTTG GATCCTTCAT CAAGAGTTCT	6480



1161

TCACTAAAGA GGGCATCATA AGGAACTTTC CCATTATAGT AATGATAGTG GTCACCGTGT	6540
GACGTTACAT AGCCCTGATC TGTAATTTTG ATTACAATTT GCTCAGCCTG AATTCCTTCT	6600
TTCTGGCTAA CCTGGTCTGG TGTCAAGTTT TCACTTTTCT GACTTGACTG GCTGCCATCC	6660
ACATAAGAGA CACGATTATT GTCCTTATTT TCCTGCGAAC GATGCTGGTT TAGTGCATAG	6720
GCACATAGAC TCAAGGATAC GATAACAGCT GATCCAGCTG CTATATATTT TTTACTAAAT	6780
TTCATAAATC CCTCATTTCA ATAAATGATG AAGTTTTTTC TCAACTTCTT TTACTTTATT	6840
AAATAGTTTT CTAAACCCGG GGGTACC	6867

(2) INFORMATION FOR SEQ ID NO: 193:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 999 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

CGTTCTAAAA ATGCAGTACG TTTGATTGAG AAATCAGTTA AAGGTATGCT TCCACACAAT	60
ACACTTGGAC GCGCTCAAGG TATGAAGTTG AAAGTATTTG TTGGAGCTGA GCACACTCAC	120
GCTGCACAAC AACCAGAAGT TCTTGACATT TCAGGACTTA TCTAAGGAAA GGAACAATAA	180
AGTATGTCAC AAGCACAATA TGCAGGTACT GGACGTCGTA AAAACGCTGT TGCACGCGTT	240
GCGCTTGTTT CAGGAACTGG TAAATCACT GTTAACAAAA AAGATGTTGA AGAGTACATC	300
CCACACGCTG ACCTTCGTCT TGTCATCAAC CAACCATTCG CAGTTACTTC AACTGTAGGT	360
TCATACGACG TTTTCGTTAA CGTTATAGGT GGTGGATACG CTGGTCAATC AGGAGCTATC	420
CGTCACGGTA TCGCTCGTGC CTTCTTCAA GTAGACCCAG ACTTCGCGA TTCATTGAAA	480
GCGCAGGAC TTCTTACACG TGA CTCACGT AAAGTTGAAC GTAAGAAACC AGGTCTTAAG	540
AAAGCTCGTA AAGCATCACA ATTTAGTAAA CGTTAATTCG AAAGAATTAC TACTATTATA	600
CAGAGCACCT TTCGGGGTGT TCTTTTTTTA TACTTTCTTA CTAAATTGGT GCAATTGACA	660
CAGTTGTGTC GACTTTAGTC GCTTACAAAT GTGGCTGCAA CCTGACATGG TCAGTTGCCT	720
CAAAACGTTA ATCAATACGA TTATATCAAC GTTCAAAGC ACTCAAGGGT TTACCCTATG	780
GGTGCTTTTT TCTATACTTT CTAAAAAGT TTACCCTAAA ATTTGCCCTA AAATTACCCT	840
ACTTATTTTT AAGATGTTGG TAGGCAACTT GTCCAGCAGA TAATGGAAC ATGTTTGAAG	900
TATTAACATA AGTCTTAGTT GTAACGGTAT CGCTATGACT TAATGCTTCA GAAATGGCTT	960

CTAAGTCAT TCCTGCTTTT TTAGCAAGTG TCGCTCCTG

1162

999

(2) INFORMATION FOR SEQ ID NO: 194:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2315 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

AATATTATCA CTGTTCTTGA AGGCAGAACA CAAGCTGTCA TCCGAAATCA CTTTCTTCGC	60
TACGATAGAG CCGTTCGTTG TCAAGTGAAA ATCATTACGA TGGATATGTT TAGTCCTTAC	120
TATGACTTGG CTAACAGCT TTTTCCGTGT GCTAAAATCG TTCTAGATCG TTTCCATATT	180
ATCCAACATC TCAGCCGTGC CATGAGTCGT TTTCGTGTTC AAATTATGAA TCAGTTTGAA	240
CGAAAATCTC ATGAATACAA GGCTATCAAG CGTTACTGGA AACTCATCCA ACAGGATAGT	300
CGTAAACTCA GCGATAAAGC TTTTATCGC CTTACTTTTC GCATGCACTT AACAAATAAA	360
GAAATTCTTG ACAAGATTTT AAGCTATTCA GAAGACTTGA AACACCACTA TCAGATCTAT	420
CAACTCTTAC TTTTCACTT TCAGAACAAA GACCCTGAGA AATTTTTCGG ACTCATTGAG	480
GACAATCTGA AGCAGGTTCA TCCTCTTTTT CAGACTGTCT TTAAAACCTT TCTAAAGAAC	540
AAAGAGAAAA TCGTCAACGC CTTTCACTA CCCTATTCAA ACGCCAAATT GGAAGCGACC	600
AATAATCTCA TCAAATAT CAAACGAAT GCCTTTGGTT TTCGAACTT TGAAAACCTC	660
AAAAAACGGA TTTTATCGC TCTGAACATC AAAAAGAAA GGACGAAATT TGTCTTTTCT	720
CAAGCTTAGC TTTTCTTCAA CCCACTACAG TTGACAAAGA GCCTATTTTC GCTGATTCTC	780
CACTACATTT GACTGGATT CTAATTTTTTA GAGAAATACA AAAGAGCTAG CTTTAGCTAG	840
CTCTTTTCCT ATGCGGAGAG AGGGACTTGA ACCCTCACGA CCTAAAGCGG TCACAGGATC	900
CTTAGTCCTG CGCGTCTGCC AATTCCGCCA TCCCCGCGTC GATTACTTTA CTAGTATATC	960
AACTTTTGGG ATGCTTGTC AACTTTTTTT TCAAATTTTT TCATTTTCAC CAACCAGGTT	1020
ACTCAAAAAG TTCATTTAGA TTTTCATCTA CTAACCTAGC TCCGAGTGTA TTTTGAAT	1080
GACCTAGGGC AAATTGATGA TTTTCAGGCC AGATGGAAGC AACAGCTGGT TTAACAATCT	1140
CGATGTCATA TCCTAGATTA TAGGCATCTA TAGCTGTATG TAGGACACAG ATATCCGTCA	1200
AGACACCTGT TAAGATAACG GTAGACACTC TACGCTCTCT CAAACGAATA TCTAGGTCAG	1260
TCCCTGAAAA AGCTGAGTAA TGGCGTTTAT CCATCCAAAA GACACGACTG TCTGAACCAT	1320
GCTCTTGATA AAAGATCCCC AAATCTCCAT ATAAATCCG TCCACTCGTC CCAATCAGAT	1380

1163

TATGAGGAGG AAATAACTTA CTTTCCGGAT GGAAACAATC GTTTTCTTCA TGAGCATCAA	1440
TAGTAAAGAA GATATAATCT CCTCGTTCAA AAGCTAATCG AGTTACCTTG CTGATGGCAT	1500
CCGAAATCGC CTGAGCTGGA GCACCTGCTG TTAGTTTCCC ACTATCAGCA ACAAATCTT	1560
CTGTATAATC AATCGAAATT AAAGCCTTTG TCATTAGTAA TCTCTTTCT TCACTTCTTC	1620
AAAAATATCT GAAATCAAGA CCTTAAGATA GGTTCCTTC ATTCCAAGTG AGCGACTTTC	1680
AATAATCCCC GCAGACTCAA GTTTACGAAG AGCATTGACA ATCACAGAGC GAGTGATTCC	1740
GATACGATCT GCAATCACTG ACGCAGTCAA CTTCCCTTCA TTTCCATTTA ATTCCCCTAA	1800
AATTGCTGAA ACAGCACGGA GTTCGGAGTA AGAAAGGTA TTGACCGCCA TGGTGACAGC	1860
AGTACGACGA CGAATATTTT TCTCATCTTC TTCACGTTGG AAGTTAAGAA GCTGAATCCC	1920
AACAACGGTA CTGGCAATCT CAACAAGAAC CAAGTCCTCA TCTTCGAATT TTTTATCATT	1980
ACGCCAAATA ATCAAAGAAC CAAGGCGAAT CCCCATACA TGAATCGGTG CAATAGTCGT	2040
CAAGCCATCT GGAAATCAT CTCTACTCTC AATAGGGAAA ATACTCATAT CATGCTCAAC	2100
AGGCAAGTTT GCTTCTGTTT CGTAAATCAT ATTAGCCCTT TGAACGTAGT CATCTGGGA	2160
AATCTTAGTT TGGAAGAATT GCTACGCGA TCTGTATTG TTTTATAACG CATAAATAG	2220
CCAAGCAGAC GTCCCTTACT ATTGATAATG CAGGCATTGC AATGAATAAT ATCCGCTAAC	2280
TGACGCGTAA TAGCGTTGTA AGGGAGCTCA TCTCG	2315

## (2) INFORMATION FOR SEQ ID NO: 195:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6693 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

CGATTTCTTC CATTTCTTCA AATAAGAATA CTTATCTGA CATATGTGTT ACCTTCTTCA	60
TCAAAAATTA TTTTGTAATC GATTACATTG CAGATCGTAA CATAAAGAAA AACAGATGTC	120
AAATATTAAA CGTAAAAACA TGGTCACTAA AGAACTATAA GAGAAAAGGT AAACCTAGCG	180
ACCGGATGAA CGCTGGGTCG TTTGGTTTCG ATTGCTCTCT TCCTCTTGTT TTTTCTGTTT	240
TTCTTCTTGT TTTTCTCAG CTTCTTGGC CTCTGTTTG GCTTTTTCCT CAGCTTCCAT	300
AATTAATTTA TCCGCCACAG TGAGCTGTA GATTCCAGCT TCCATGTCGA CCACACTCGG	360
TTCTGACAAT TGAGGCTTAA TCTTACTGTA ATATGGCAGT TTCTTACTCA TTTCAGATAG	420

1164

AGGAACCAAG ACTTCGTCCG AATCATTTCAT GGTCAATCGA ATTAAATCGG ATGTCACCTT	480
GCTTGGGGCT AATTCCACCT TTTGGATAGC CGCCTTGAGT TCTGGGCTAA TTTGAGCAAG	540
TTCTGAGACA AAAACTTTGA TTTGTTCACT ATCATTAAAG AGAACTGATA AATAAGTTTC	600
TGGTAAACTG TTCAGACTCA CAGAACTAGT CTCAAGCTGA CCACTGGAAA GAATAGGATA	660
ATGATTTTCA CCAGAAATAT AGTAGGCCAC AATATCATAT TCCTTGACCT TAATAGTGAA	720
CTTAGTTGGA AATTGATAGA CAAGTTGAGC TGATTCAACC CAATAGTTAG ACTTAATCTG	780
CTTTTCATAT TTTGCCTTGT CTAGCAGAAG GTTAATCGTA TAATCCGAAT CCTGAATGCC	840
TGAAGCCTGT CGAATATCAT CAGCTGTAGT TTGCACCGTT CCCTCAACAC GAATATCTTT	900
CATGGTCGCA TAAGGACTGA GCAAGTAGGC AGAGACAAAC AATAGAAGCA GACTTGAAAA	960
TAAAATCGTG AAGGCTCGCA AGATATGGAT ACCAGGAATC TTTGCTTTGG CTGGTTTTTC	1020
CTTTGTAGCC TTTTAGCAA GCTTTTTATC CTGTTCTCC TTCTCTTTAG ACTCTGGTTC	1080
TTCTTTCTCT TCTTTCTCTT TGTACGCTC TGAGGATGCT ACTTTTTCTT CAGACTCTTC	1140
CTTAGCTGAT TCTGAATCTT CCTGGTCTGT TTCACTCTCC TGGTCCTGTT TATCCTCTGA	1200
CTTCTCAGAT TCTTCTCCCA TCGAGCTTG TCTTCTCTT TCCTTCTCCT CAGCTAGAGC	1260
CGCCTCTTCT TCAGCCTTCT TTTTAGATA TTCTTGGTTT CGTTTCTGCC ATTCTGATAA	1320
CTCTTTCAAT TCTTCGAGGG TTTCTTTGTC CTCATTTTTC TTATCTTTTG ACATTTACTT	1380
TCCTTATGAT AAATCTTTTT TCAACAATG AAAAAATCT GCTAGAGATT TCAATTCCTT	1440
AGAAGCTTTC ATCTTAGCTT GGTAATCTTC CTGTGTACTT AGTAAGTGAG AAAGCTTCTC	1500
TTCCAAACTA TCCAAGGTCA AATCGCTTTC TTGAAGGTCT TCTGCATAGC CTTTCTTAAC	1560
AAAGTAAGCT GCATTTTCAA TCTGGTCACC ACGACTAGCT TCACGACCAA GCGGCACAAT	1620
GACATGCAAT TTTGCTATCG CCAAGAGCTC AAAAATCGTA TTGGCACCAC CTCGTGTCAC	1680
AACAATATCA GCCAATTCCA TCAAGGGTTG ATAGAGATCG GTCACATAGT CAACACGAAA	1740
AAGATTTTGC CTCAACTCAT TCAGACTAGA ATCTCCAGTT AGATTGATAA TATTGTAGCG	1800
CTCTGTTAGT TCTTCTTAT GGTCTGTCAC CAATTGGTTA AAGACACGAG CGCCTGCAGA	1860
ACCGCCAACA AACAATACAG TTGGCAATTT GGGATTAAAG TGGGTTTGAA TATCCACCAA	1920
TTCATCTGGT TCTGGAGTGT TTTGTCCGA AACCTTGGTC ACCGCTCCCA CATGCTCAAC	1980
CTTAGCCAAA CTCGAAGCTT GTTCAAAGGT TGAATACATC TTAGTCGCAA ATTTATAGGC	2040
GATTTTATTG GCCAAGCCCA TAGACAGGTC AGATTCGTGA ATAAAGACAG GCACTCCTGA	2100
CACACGCGCA GCGATAACAG GCGGTACTGA GACAAAGCCC CCCTTTGAAA AAAGGGTCTG	2160
TGGACGCGT CGCAACATGA TAAAGAGCGA TTGGACAATT CCCCACCAA CTTTGAAGAC	2220

1165

GTCCAGCATA TTTTGCCAAG AGAAATAGCG ACGCAATTTT CCAGTCGCAA TAGAATGGAA	2280
GGTGACATCC AAACCTGACT TAAGGATTTC TTGGTGTTCG ATACCACACT TGTCCCGCAT	2340
ATAGTGGACT TCCCAACCAT CTTGATGAA CTGGGCATT AACAAAAGAT TGAGGGTAAC	2400
GTGTCCAACC GTCCCCCAC CTGTAAAGAC AATTTTTTTC ATATTATTCT TTAACTCCG	2460
CTACTGTGTC GATAAAGAGG TCGCCACGTA CTTCAAAGTT AGCATACATA TCCCAGCTAG	2520
CATTGGCAGG ACTAAGAAGA ACCACATCTC CTTGAGTCGC AAGCTCATAG GCCTTGCGGG	2580
TCGCATCTGC AATATCTGTC GCCTCCACAT AAGCGACACC AGCCTTGTCT GCTGCCGTT	2640
TGACACGTTT TGCAGATTGA CCCAGGATGA CCATCTTCTT GAGTCCAGTA ATGTCTGGCA	2700
CCAATTCGTC AAACCTCATG CCACGGTCCA AACCACCTGC AATCAAGACG ACCTTGCTGT	2760
TGTCAAATCC TGACAAGGCT TTTTGAGTAG CCAAGATATT AGTTGATTTA CTGTCGTTAT	2820
AGAATTTAAC ACCCTTGATG TCATCCACAA ACTGGAGACG GTGTTTGACA CCACCGAAGG	2880
CTGAAAGAGT TTCCTTGATG GTTGATTGT CCACATCAGC AAGCTTGGCT ACAGCAATAG	2940
TCGCAAGGGC ATTTTCCACA TTGTGGCTAC CTGGAACACC GATTTTCATC GCTGCCATGA	3000
CTACTTCACC ACGGAAGTAG AGTTGACCAT CTTCCAGATA AGCTCCATCA ACCTTTTCAA	3060
GTGTTGAAAA TGGTACAACA GTGGCTCTG TCTTGGAACT CAAGTCTTTT GCCAAGTCTT	3120
GATTAAAGTT CAAGACAAGG AAATCAGCTG CTGTCATCTT GTTCTGGATA TTCCACTTGG	3180
CTGCTACATA TTCCGAAAAT GACCCATGGT AGTCGATATG AGTTGGCATG AGGTGGTAA	3240
TAACCGCAAT CTCTGGATGG AATTCCTGAA CACCCATGAG TTGGAAAGAA GAAAGTTCCA	3300
TAACAAGCGT GTCCTTATCT GATGCTATTT GAGCAACCTG ACTAGCTGGA TAGCCGATAT	3360
TCCCTGATAA AAGACCATGT TGGCCAGCAG CAGTCAAAAC TTCCCAATC ATAGTCGTTG	3420
TGGTTGTCTT ACCGTTGAT CCTGTGATAC CAATAATCGG TGCTTCTGAA ATCAAATAAG	3480
CCAATTCAC CTCAGTCAAG ACTGGAATC CCTTGGCCAA AGCCTTTTCA ATCATGGGAT	3540
TGTTGTAGGG GATACCTGGA TTTTACCA TAAGGGCAA CTCTTCATCC AAGAGTTCCA	3600
AAGGATGGCC ACCTGTAATG ACCTTGATCC CTTCTTCCAG CAACTTTTG GCAGCTGGAT	3660
TGTCCTCGAA AGGTTTCCCA TCATTTACTG TCACAAATGGC ACCTAGCTTG TCCAACAAAC	3720
GAGCTGCAGA TTCACCAGAC TTGGCCAAAC CTAACAAG GACTTTCTTA TTTTAAATT	3780
GATCTATTAC TTTCATGTCT CGAACTCCAT TTCTACTCCT ACTATTTTAC CATTTTATG	3840
GAAATAAAAA AGCCACAAAG TGTGTTGTG ACTCTTTCTT CTAAGTGAAT CTTACCATAT	3900
CATCTATGTG ATAAATCGGT AACTCGAATG ACCTGATCCA CTTGCTCCCA AATCAGAGGA	3960

1166

TTATGGGTCG CAATAATAAT GGTCCGATTC GGATTTTSTA AAGATTCTAG GATGGAAAGT	4020
AATTCCTCAG AGTTTTTGGG GTCTAAGGAA GCGGTTGGTT CATCTGCCAG GATCAAAGGT	4080
GGATCCTTTA AAATTATCTT CGCTAGTGCA ACACGTTGTG CTTCTCCTCC TGATAACTCA	4140
AATATAGGTT GCTTCAAATC CAAATAAGAG AGGTTTACAC GGTTTAGAGC TTGTTTCATC	4200
AAAGAGATTT TCTCTTTTTC CTTCAACTTT TTACCAACTA AACCCAGATT GAGATTCTCT	4260
TTGACGGTTT GGCTTTCAAT TAAGCCAAAA TCTTGAAATA AGTATCCTAA GTAATCTCTA	4320
AAGAAAACAG AAGGCTTGAT GTCCTTAAGA GAAGTGCCAT CATAGATGAT TTGCCCTTTG	4380
TCATATGGCT CCAATCGTCC AATCATATTC AAGAGTGTG TCTTACCACA GCCACTTGTA	4440
CCGATTAAGG CATAAATTTT CCCACCTTCA AAATGAAGAT TCATATCTGA AAATAGCTGA	4500
CGGCTTCCAA ATTTTTTAGA TATATTCTTT AGTTCAATCA TCCTATTTTC CTTTCATAAT	4560
TGTCATAGAA ACACGAGATT CTTTCTGCGC TTGACGGTAA AGCGTCAAAA CTGCACTAGC	4620
TAGAAAGACC AATAAAGTGA GCAAGCCAAT CACCAAGTCT CGACTGCTTA AAATAAAGAG	4680
ACTAGCACCA AATACAAAAC TAGCAAATG GCTAACCATA TACTGAGCAT GTGTTTCAAA	4740
AAATCGTAAA CCTGAAATTC GTTTAATCAA GATATCTCGG CGGAATTGCT CGAAATATAG	4800
AAGATTGACA GAATAAAGA GTAACAAGGA ACTGGCTATT CCAACAATAG CTCCTAAGAT	4860
TAAAGTTGCT GTTTCAGTTT GAACCTCATT ATAACGAGTT AGATAAACAC TTCTTCCTTC	4920
TTTAAGATAG GATACTTGCT CATAAATTC AGCTTTCTTC AAGAGTTCTA GCCCCTCTC	4980
ATATCCTTTG ATAAAGAGTT GTTTTCCAGC ATTGATAGAC CAACTAGATA AGGATATAAA	5040
ACTATCACCT GTAGAAGTCG GCGTGAATAC CACTAAAATC GGATCAGTCA AATACTGAGT	5100
AGATACGGGA TTCTCACCGT TATTATAAAC AAACCGCTTT TCTCCATTG AAAGATAACT	5160
AACGTGCGCT TTCTCTCAT AATCCAAAGG AGCACTTGCC TCCTCACCAG ATTTTCCATA	5220
ATAACTCAAT CTTTCTTCAA AAACCTTCTT AAGTTCTGCT TCTCGAGAGC GCAAATGTTT	5280
TGGGAGCAAG AGGATAAACT CACCTTTTGT GAGATGGGCT AACTTCTGTT TGGTCTCAGC	5340
ATCTACCAGC ACCTTTTCCT TGTCCAAATA ACTGGGACTA ACATAGAGCG TATTAGCATC	5400
TGAACTATAG GTATCCAGTG TCTCTCCCTG TTCATTTTTT CTTGTGGAT TGGCAAATG	5460
GAGCAGATTA TCCTTTACAT AAAGAGCTTG TTCTTCTTCG ATTGCTTCCT TGGCAAAGGC	5520
ATACCACTTG CTCTGATTTT CTGTATCTTT TCCTCTATCA CCTAAGCCAA AGGAAATCTG	5580
GTAATAGTCT GCTCTGTCTT GCCATGCTTG TTTTGAAATT TCAAGTTCTT TCAATCGTTG	5640
GTAAGACGTC AAACCTGTCT TAACAGCGTA GCCTACTGTA AAAACAGCTA CTAAGTACCA	5700
CAATAGGGTT AAAGCCATCA AGCGTTTAAG GGGTAATCTT CCCTTAATAA CGGGAACATA	5760

1167

TGCTTTGTAA CTCAAACCTCA TTAGGTAAAG GAGCATTAGT AAAATTGAAA TCGCCAATAA	5820
AAACAACAGA TAGAAACTAA TCCCAAAACC ATAGGTGGCT AACAAGATAG GATAAAACAA	5880
ACCTTGACTA AAAAGAACGA CTCCCCCACC TAGGAAGGAA AGGAGGGCTG ATAGAAGGAG	5940
CCATTTGATA TCACTAGATA AAGAATGCCC CATGATGGAT AAGAGAGTCT GACCAGAAAA	6000
GAGTTTTATA CCTGCTGCTC TCATTTCCCTT AATCCGAGTG ATAATCACTA AAGCAAAGAA	6060
AGATAAGCCA AATATTGCTA AACTAATTAA AATAAGGGGA TTTAGTAATA TTCGAAAAGC	6120
AAGAAAATAG GGCGGTATCT TTCGGTCAGC ACTTGCTTTA TAACCCAAAT CTCCTAATTT	6180
ATCGGCAAGC TTTTCTTTTCG TCAAGGAGCC TGACAAAAGG AGATAACTAT TTAGCGGANT	6240
ATACGTTTAC GACTTTCTTG GCTAGCTTCT TGGAATTCTT TTGGTAAAGT TCCCTGACCA	6300
TAAGTTGCAT AAGTAAAGTG AGTCGTCCCA TCCTTACTCG GCTCTACAAT TCTTCTAGCT	6360
ATTAACTCT GTTCTGAGTT TGCAAAATTC TCCAATTCCT GTTCAAATAC CTCACGCGTC	6420
GGTTCCTGAG TATCTTTTTT GACACGAAGT AAAGAAACGG AATCATAGCT TGATATAAA	6480
TATTGTGGCG CACGTAAGAC AATAATCCAA GCAAGGAAGA AGCTGAGAAA AAAAGTTGAT	6540
AATAATATGA ATAGTTTCTT CATAGTAGAC TCCTTGTAAG CAAAATTCCC CCTGTAATTT	6600
CTTACAAGGG GAACGATTTA AATCAATGAA CGATTAGTCA TAATCACAGT AAAATGCTAC	6660
TTGTTCTCCC CATTTAGTCC AAATCCATGC AGG	6693

(2) INFORMATION FOR SEQ ID NO: 196:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1847 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:

CCGGTCTATG TACCCACTAC TTTGGGACAA TATGGGGATC AGCTACCCAA AACTAATCGA	60
GCGTTTGGTT GACCTTGCCA AGGAAAGTTT TGACAAGCGC GACGATTTGA TATAAATGA	120
AAGAGAGGGT AGAAGCCAGA ACCATCACTG CACGGTGACT AGAGTTCTCG GACTTCAGCC	180
CTTTTAAAG GAGTAGAAAT GAAATTAACA ATCCATGAAA TTGCCCAAGT TGTGAGGCC	240
AAAAATGATA TCAGTATCTT TGAGGACACC CAGTTAGAAA AAGCTGAGTT TGATAGTCGT	300
TTGATTGGAA CTGGAGATTT ATTTGTGCCA CTAAAGGTG CGCGTGATGG CCATGACTTT	360
ATTGAAACAG CCTTTGAAAA TGGTGCAGCA GTAACCTTGT CTGAGAAAGA GGTCTCAAAT	420

1168

CATCCTTACA TTCTAGTAGA TGATGTTTTG ACAGCCTTTC AATCCTTAGC ATCCTACTAT	480
CTTGAAAAAA CGACTGTTGA TGTCTTTGCT GTTACAGGTT CAAATGGCAA GACAACGACT	540
AAGGATATGT TGGCGCATTT ACTGTCAACA AGATACAAGA CCTACAAAAC ACAAGGCAAT	600
TACAATAATG AGATTGGCCT TCCTTACACA GTTCTTCATA TGCCTGAAGG AACAGAAAAAG	660
TTGGTTTTGG AGATGGGACA GGATCACTTG GCGGATATTC ATCTCTTGTC TGAATTGGCT	720
CGTCCAAAAA CAGCCATCGT GACCTTGGTT GGAGAAGCCC ATTTGGCCTT TTTCAAAGAC	780
CGTTCAGAGA TTGCTAAGGG AAAAAAGCAA ATTGCAGACG GAATGGCTTC AGGTTCTCTG	840
CTTTTAGCGC CGGCTGACCC TATCGTAGAG GACTATTTGC CAACTGATAA AAAGGTGGTT	900
CGTTTTGGGC AAGGGGCAGA GCTGGAAATT ACTGACTTGG TTGAGCGCAA AGATAGTCTG	960
ACCTTCAAGG CCAATTTCTT AGAGCAAGCC CTTGATTTGC CAGTAAGTGG CAAGTACAAT	1020
GCGACAAATG CTATGATTGC ATCCTATGTT GCCTTGCAAG AAGGAGTTTC AGAGGAGCAA	1080
ATTCGTTTGG CCTTCCAAGA TCTTGAATTG ACGCGTAACC GTACCGAGTG GAAGAAAGCA	1140
GCCAATGGAG CAGATATCCT GTCAGATGTT TACAATGCCA ATCCAAGTGC TATGAACTG	1200
ATTTTAGAGA CTTTCTCTGC CATTCCAGCC AATGAAGGTG GCAAGAAAAT TGCAGTGTG	1260
GCGGATATGA AGGAGCTTGG TGACCACTCT GTTCAACTTC ATAATCAGAT GATTTTGAGC	1320
CTTTCTCCAG ATGTGCTTGA TACCGTGATT TTCTATGGAG AAAATATTGC TGAATTAGCC	1380
CAATTGGCCA GTCAAATGTT CCCAATCGGC CACGTTTACT ACTTCAAGAA AACAGAAGAC	1440
CAGGATCAAT TTGAAGACCT AGTCAAGCAG GTCAAGGAAA GCCTTGGAGC CCATGACCAA	1500
ATCCTGCTCA AAGGCTCTAA CTCTATGAAT CTAGCCAAGT TGGTAGAAAG TTTAGAAAAT	1560
GAAGACAAGT GATTTTGTCA AGTATTTGCA AAGAATGATT GCCATTACAG ATACTGGCTT	1620
AACCTTTACA AAAGATCCGT TTGACCGTGA GCGCTACGAA GACTTGCGAA GTCTGTTATC	1680
TGAAATGTTG AATCAAGCAT CAGACCTTGA TTCCGAAGAA GTGGCAGAAG TCTTGAAGCC	1740
AACCTTCTGCT TATGCGACTC CGTTAATGGA CGTCCGTGCT TGGATTGTTG AGGATGAGAA	1800
GATTTGTCTG GTTAGGGGAC AAGGAGAGGA TAGTTGGGCT TTGCCGG	1847

(2) INFORMATION FOR SEQ ID NO: 197:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1062 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:



1169

CAAGCGAAAA CATT TTTTAT TCCAAATAAA CAGAGCATT TAGGAGAACA AGAGATTTTG	60
AATGCCAAGT CGATCTTGGC CTTGCTAGAC GGT TTTGAGT CACATAGCTA TGATGTAGTC	120
TATCTCCGTC AGCCTCTTAA TCGTCTCGAA TATATCGAGT GTGCGATAGT GGGGCAATCA	180
CAATTTCTCT TTAAGGTCAG TTATGCTGAT GGTCAAAAGC CTTACCGTGT CGATCTTCCT	240
GACCTACTAA CAAAGACAGA CTGGCAGATT ATCAAGTCAT TTTTAGATGC TTTGCTTGCT	300
TATACAGGGA CTGATATTGA AGGGCTAGAT GGT TTTGATT TTGAAGCTTA TTTCCAAGCA	360
AGTATTCAAG CCTATCTAGC AGACCCTGTA GCTCGTTTTA CGATTTGCCA AGGAATTTT	420
AATCCTATTT TCTTTAGTCG TGAGAACTTG AAAAGCTTTT TAGAGGCAGA TGGCTTGCT	480
CAGTTTGAAG CGCGTGTGCG TGCGGTTCAA GAGACAGATG CCTACTTTGC GAGAGTTTCC	540
TTCTATCAGG ATGGAGAAGG AAAAGTGCAT GGC GTTTACC ATCTAGCTCA AGGAGTCAAG	600
ACAGTTTAC CGAGAGAACC GTTTGTTCTT GCAGCCTATA TTGAGCAATT GGTGGATAAG	660
GAAGTCCAGT GGGAGATTGA CTTGGTTCAA ATCAGAGGAG ATGGCTCTAA ACCAGAAGAC	720
TATGAAGCCA TTGCTCGCTT GGA CTATGCA AAATCTTAG AGGTATTACC CCCATCTTTT	780
TACCACCAAC TAGACGCCAA TCAAATAGAA GTGCAACCCA TATTAGACAA AGATTTTAAA	840
ACATTAGCAC AAGAAAAGTA AAGCAGAAGC AGGTCAATCG ACTTGCTTTT TTGACATAGA	900
AAAAATCCTG CCAAGATGAC AGGATTGCTA CTCAATGAAA ATCAAAGAGC AAAGTAGGAA	960
GCTAGCCGCA GCTGTACTTG AGTACGGTAA GGCGAAGCTG ACGTGGTTTG AATTGATTT	1020
TTGAAGAGTA TGAAGTTTAA AGAAAAGCCA AGATACGAAG AT	1062

## (2) INFORMATION FOR SEQ ID NO: 198:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6846 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

TATCTACAAC CTCAAAAACA TGTTTTGawG gCTCGTCAGT cTATCTACAA CCTCAAAAAC	60
ATGTTTTgAa kGctcGTCAG tTCTATCTAC AACCTCAAAA ACATGTTTTG AcaGCcTcGT	120
CAGTTCTATC TACAACCTCA AAAACATGTT TTGAGCTGAC TTCGTTAGTT TCATCTACAA	180
CCTCAAAAAC ATGTTTTGAG CTGACTTCGT TAGTTTCATC TACAACCTCA AAAACATGTT	240
TTGangnCnT CGTCAGTTCT ATCTGCAACC TCAAAGCAGT GCTTTgagcG CTTGTCAGT	300

1170

TCTATCTACA	ACCTCAAAAC	AGTGTGTTGC	GCAGCCTTTA	ATCAGCCGCC	TAGTCCGCTC	360
TATGGTATTC	ATTAAGTCAA	CATCTCTTGT	TTAAGAGCAC	CAAATCAGGA	AATCTTCTCG	420
ATTCCCTGAT	TTTTTCTATT	TACGTTTTTCG	TGTTGAGCTA	CGTTCTGTCA	AACCATGAGG	480
TAAGAGAACT	TCACGTTCTT	CCAACTCTTC	CTTATGCATA	ATCTTGGTCA	ACATACGCAT	540
ACTAATGGCA	CCAAGGTCAT	AAAGAGGTTG	GGCAATCGTT	GTCAAGTTTG	GACGGGTAAA	600
GCGTGAGATT	TGTGAATCAT	CACTAGTAAT	AATTTCAAAA	TCTTCTGGCA	CAGAAACACC	660
CTTATCAGCC	AAACCGTTCA	AGACTCCTGC	TGCCAACTCA	TCACCTGTCA	CAACTGCTGC	720
AGTTGCATTT	GATGAAATCA	AACGCTCTGC	TAAGGCGTAA	CCATCATCAT	AGCTATATTT	780
AGATTCAAAT	ACCAAACCCCT	CACTATAAGT	GATTCTGCT	TTTTTCAAGG	TTTCCTTGTA	840
GCCAACTAAA	CGAACCTTAC	CATTGATGTC	ATCCACTAGC	GGACCGCTAA	CGAAAGCAAT	900
ACGCTCATTT	TCTTTAGCAA	GGTAACTCAC	TGCATCAATT	GTTGCTTGCT	TATAGTCAAT	960
ATTGACACTT	GGCAACTGGT	GCTCAACATC	GACAGTTCCT	GCGAGAACAA	TCGGAGTACG	1020
TGAACGCGAA	AATTCTGAGC	GAATTTTATC	TGTCAAGTGA	TACCCCATAT	AGATAATGCC	1080
ATCTACCTGC	TTTGAAAAGA	GGGTATTGAC	AACAGAAACT	TCTTTCTCGT	TATCTTCATC	1140
GCTATTAGCT	AGGACAATAT	TGTACTTGTA	CATTTCTGCA	ATATCATCAA	TCCCCTTAGC	1200
CAAACTCGAA	AAATAACCAT	TGGTAATATT	TGGAATCACG	ACACCGACAG	TGGTTGTCTT	1260
TTTACTTGCA	AGACCACGCG	CAACTGCATT	TGGACGATAA	TCCAAACGAT	CAATTACCTC	1320
TAGCACTTTT	TTACGGGTAT	TCTCTTTTAC	ATTTTATTG	CCATTGACCA	CACGGCTGAC	1380
CGTCGCCATG	GAAACACCTG	CTTCACGAGC	GACATCATAA	ATGGTTACTG	TATCATCTGC	1440
ATTCATTCTT	TTTCCTGTCC	TTTCTATCTC	ACACATTCTT	TTACAAGTAG	AGGTAAGTAT	1500
TGAAGCTCTA	TATCTACTTA	CAAAAGTGAA	GATGTGAAAA	TTTCGTTTTT	ATATTTCTAC	1560
TTATTCCATT	CTATCACTAA	TTGTAAACAC	TTTCAAGTGT	TTTTTGAAGA	TTGATTGAAA	1620
AAATTTTATA	GAAAACCTAG	GTTTAGCTCC	TTGCTACCAC	CTTAGACTAA	ACAAAAAGGA	1680
GGAAACTAAG	CCCTCCTAAA	GTTATAGTAA	AATGAAATAA	GAACAGGATA	AATCGATCAG	1740
GACAGTCAAA	TCGATTCTTA	ACAATGTTTT	AGAAGTAGAG	GTGTACTATT	CTAGTTTCAA	1800
TCTACTATAG	GTATTGTTCC	ATTCACTACC	GTCAATTTTA	GCACATAGTC	TTCATGAAAA	1860
TATTATATCA	TCATAACCAA	CCAGATTCTT	TCGCGATATT	AGCTGCCTCT	GTTGATTAC	1920
CTGCATCTAG	TTTCGAAAGA	ATATTGGTGA	CATAGTTTCG	GACTGTTCCG	TTGGATAGAT	1980
AAAGTTTGTC	TGCAATTTCT	TGGTTAGAGA	AGCCCTGAGC	AATTCCTTTT	AAAAGTGGCA	2040
TTTCTTGCTC	CGTTAATGGA	TTGGGATGCA	TCATCACCAC	TTCCATCAAT	TCAGGCGAAT	2100

1171

ACTCCTTGCG TCCTTCGAGG ACGGTGTGCA AGGTTTGCAT GAGGTCTGCA ATGTTTCTTT	2160
CTTTTAATAC ATAAGCATCT ACTCCAGCCT TGACCGCAGC TTCAAAATAC CCAGGACGCT	2220
TGAAGGTCGT CACCACAACC ACCTTTGTTT CAAGCTTTTC TGCTCGTATC CACTCCAAGA	2280
CTTCAAGACC TGTCTTAACA GGCATTTCTA CGTCAAGGAT GCGGATATCT ACAGACTCCT	2340
TTTCTAATAG TTGGATTGCT TCTTGCCCAT TCTTGGCTTG AAAGACAGAC TCTACATCCG	2400
GTTGAAGCAT GAGCAACTGG CACATGGCAT CTCGCAACAT ACTTTGATCT TCTGCGACTA	2460
ATACTTTCAT CTACTTCTC TCCTTATAAA GTAGTCGAAC CTGCACTTCA GTTGGATGTT	2520
TCTGACTGAT TACACTTACT TCTCCTGAAA ATGCAAAAAC ACGATTTCGG ACTGTATGGA	2580
GCTCATCCCC GCTTATAGAG GCAAAGCCAC AGCCATCATC TCTCACTGTT AGAATGAGTT	2640
CTTTCTCTGT CCGTTCTAAT TTCAAGTAGA CTTTAGACGC TTTAGCATGT TTGATGATAT	2700
TGGTCACTAA TTCAAGCAAA ATCATGGAAG CCGTTGACTC CAATTCCTGA GTTAAGCTAG	2760
ACTTGTCCTCA GTGATTCTCA ACTTGAACCT CAATTCAGC AATTTCTAAC ATCTTTTCTA	2820
CAGTCTCTAG TTCGGATGTC AAAGTTCTAG ACTTAAGATT TTCCACAATG GTTCGCACTT	2880
CATTCATGGA TCCTTGCTGA TCTGGTGAAT TTCTTTTAAT TCCTTTTCCA CCTGTGGATA	2940
AGCCTCCATC TGAAATAACT GCAAGGCTAA ATCTGTCTTG ACACTCAGCA TAGCAAAGGT	3000
ATGTCCCAGA CTATCATGCA AATCCTGACC GATACGACTA CGTTCATTTT CAGCAAGCAA	3060
TAGATTTATC TGAGCATTTT GCTTGACCTG AGCTTCTTTC AAATCCTCGA CAATACGAAT	3120
CCGAACCAAT CCAAAAGTCA TTAAATCGAC AAAAGTAAGA ATTACAAGTA GATAGAATAG	3180
AAACTCAACT TCGATTCTCT GAAAAATCAA CAGTTGCCCC ACAACAAGGA CTTGAGCAAG	3240
AAGAAAAGTC CAGACATGTA AAGACTTTAA ACTACGTACG CTGAAATGAT AACTTAAGAG	3300
ATTGGATAGG AAAAAGAAAA ACCAGATATA ATTAACAGCA ACAAAGGCAG TATTCCCAAC	3360
TACATAAGTC AGCATGAGGC CCCAATATAG CCAAGATAGG CGCTGGCTCT TAGTTGTTAA	3420
AACACCCAAA TATGCCACTA CAAATAGAAT ATCAATCAAT AAATGCCAGG CAGAAAGCCA	3480
CCCAGTCACT ACAGACAGGA TGGGGAAAAA CATAAAAAAT AACTGATCC AAAACATATA	3540
ATGTATTCTT TTCAGTCTTT CAAGCATTA AATTCCTCCT TATGACCTTG AAGGTAAATG	3600
GTCAAACCAA ACAAACACTAC TGAAAAACA AGTAAATAAA CTGTGGCTGA TAGATTGATG	3660
CCACCCTCAT TTAAGAAGGT CTTGAGCAAC TCCATCAACT GATAGGTCGG GAGACACTTA	3720
CCTACTACTT GCATCCAGTC TGGAAATAAA GAGATAGGCA TCCAGAGTCC ACCTAAAACA	3780
GCCAACCCTA GATAAAGAAG ATTGCCCACG ACAGACATCA ACTGACTAGT TGGTAAGAGA	3840

1172

GTCAAGGTCA AACCAAGCGC TACGAAGGCA ATACTTCCTA CTATCAGCAA AAGTGCAGCC	3900
CCAATCCAAT TTCCAAGAGA CATGTCCACA CCTCTTACAA AATGCCCAAC TGAGAAAACC	3960
ACCAAGATTG AAACCAAATA ATCAACCAGC ATACTTGTTA TCTTTGATAG ATAATATTCT	4020
ACCATATTTA CAGGGCTATG ACGCAATGTT TTCTGCCAGT TGTGATCTT GTCGGTATGT	4080
AAAACAACCTG GGAATGAGAA GATAGCTGTT GACATCATGG AAAATGCAGT CATGGAGATA	4140
AGATAATCAC GCATAAAATT CGCGAGTTCA CCTGGTGTGT CCTGATAGAT ACCAGAAAAA	4200
AATAAATAGA AAGCCGTCGG CATCCCTACT GACAATAGAT AATAGATCAA TTGTCGTTTG	4260
GTCAATAAAA ATTCTATCTT ACTAAGTGCT AGCCATCGTT TCATCTTAGT TATCTCCCTT	4320
CTGCGTTTCT TCAAAGATTG TATCCAACAA ACTACGATTA TTAACITCAA TTTCTTGAT	4380
GCCACATCCT GCTTGAACCTA ACAGTTCCCA AAAAGCATCT GCTTCGCGTG TGACTACTTG	4440
TAGAGCATCC TGTTTTTGTG ACCAGTTTTC AACCAAGTTA GACTGCTCAA TGACTTCCTT	4500
GTATGCCAGA GGAAGGATAA AATGCTTTTC AATTCCCTCA CTACGCATAG CTAGAGGCGT	4560
CGTATCAGCA ATCAACTCTC CCTTATTAA AACCAAAATC CGGTCAGCCG TATGCTCTAC	4620
CTCTTCAATA TAATGAGACG AATAGAGAAT CGTGACTCCT TGCGCTTTTA GGTCCCGAAC	4680
GATTTCCCAA AAGCGTTGAC GAGTTGAAGT ATCCATGGCA GCAGTTGGTT CATCTAAAAA	4740
GACAAGCTTT GGTGCGCCAA TCAAGGTCAA GACAAAAGAG AAGAGACGCT TTTGCCCGCC	4800
TGACAATTTT TCTGCGAATT GCTCTTTTGT TGCTGGTCA AACTGCAATA GTTGATCGAT	4860
TTCTTGATCG CTCAAGGAAT TTGGATAGAT ACGTTGAAAG AAAGCAATCA ACTCTTTGAC	4920
CTTTAATTTT TGAACGATGA CATTTTCTTG AGGCAGATAA CCTCTAATAT AGTCTAACTG	4980
AGAACTCGTC ACTGACAAGC CTGGGATGGA TACTTGACCG CTTGTGACCA GTTTATCTCC	5040
AAGCAGACAG TCCAAGAGTG TGGTCTTCCC AGCACCATTG GGCCCAATCA AGGCGACGCA	5100
TTACCTTCA GCTACCTCAA AGGAAATACC CTTCAAAATA GCCTTGCCCT TGATGTTTTT	5160
ATTTAGGCTT TCTACCTTAA TCATATTCAT GATATTCTCC TTTCAACCAC TCCATTCTCA	5220
TAAGGAAAAC GACGAAAATC ATAAATCCAA ACCCCAAAGC ACCACGAATG AATTGGCGAA	5280
GCAAGGTTTG GTCAAACCAA CCTGTAAACA TTTCCACTAA CCATACCAAG AGTGACAGGC	5340
CGATAAAGAA ATAGATGATC CCTCTCTTCA TTCTCAAGC TCCTTTTTCA CATCTCCGAC	5400
TAATTTCAAA CCTTCTCTAA CAAGCCAAGA CATCATTTCA AAGCCAGCAA AGAGCTCCCA	5460
AGGAAAATGA TAGAACTCT CATCCAATCC CGAAAACATG AGTTAGGTCA TAACTCCTGC	5520
TACTACTAAA CTCACTGCGA TAATCATTTT ATTTCTCATC TCTTCTTCT CCATTTCTATA	5580
CTACAATTAT AGTCTTTTGA AATCAGAGGA GACAGAAGCT TCTGTCACTA GAAAATATGA	5640

1173

CAAATGTCAT AAAAAATTCT GTTCAAAACA AGCAAGATAC ACTATACAAT AAAACACAAT	5700
TAGAAAAATC TAAGGCAACT TCCTCAAAAG AGATATCAAA CCCAATTCAC ACCATAATGT	5760
AAACTAATAC TTATTTAAAA TCAAAAAGAG TAGAAATTTT TATCAGACAA ACACATATAT	5820
AGTGTATTGA ATCTATAACA GTAGGCCTTA AATACTAAAA TATTTCTATA AATTAATTTA	5880
ACTTTCCTGA TAGAGCTGTT CATATCTTAT TTCAATTCTC TAAATTATAC GTTGAACAAA	5940
ACCCTTCTAT TTCTTCTTA AAGATTTATA AGAGTTATAA AATCTGTTAA ATTTCAATGT	6000
GTATACCTAA ACTACGGTAT TTATTGAAAA GACTGGAGAC AAAAAGTATA CGCTGCCAAA	6060
ATGAATTACT GAAAATCAAA AAAGAGAGAA CCAAAGTAT TCCCTCTTAA TGTATATAAT	6120
ATCTAGTTTT AAAAATACAC ACTCACATAT CTCTGTAATG AATCGGGAAG ACAGGATTCTG	6180
AACCTGCGAC ACCTTGGTCC CAAACCAAGC ACTCTACCAA GCTGAGCTAC TTCCCGAGTT	6240
AAATAGAAAA ATGCACCCTA GAGGAGTCGA ACCTCTAACC GCCTGATTCTG TAGTCAGGTA	6300
CTCTATCCAG TTGAGCTAAG GGTGCTCCAT ATTATGCCGA GGACCGGAAT CGAACCGGTA	6360
CGATCGTTAC CAATCGCAGG ATTTTAAGTC CTGTGCGTCT GCCAGTTCCG CCACCCCGGC	6420
CTCTCTAAGC GAACGACGGG ATTCGAACCC GCGACCCCCA CCTTGGCAAG GTGGTGTTCT	6480
ACCACTGAAC TACGTTCCGA CTGTTTCTT CTATCTAAAA ATGCCGGCTA CATGACTTGA	6540
ACACGCGACC CTCTGATTAC AAATCAGATG CTCTACCAAC TGAGCTAAGC CGGCTCATTT	6600
GTATATCTT AATGCGGGTT AAGGGACTTG AACCCCCACG CCGTTAAGCG CCAGATCCTA	6660
AATCTGGTGC GTCTGCCAAT TCCGCCAAAC CCGCATATAT GACCCGTAAT GGGCTCGAAC	6720
CAGTGACCCA TTGATTAAAA GTCAATTGCT CTACCAACTG AGCTAACGAG TCTAAAATAA	6780
CTTGCGTTAC CTTAAACGGT CCCGACGGGA ATCGAACCCG CGATCTCGCC GTGACAAGGC	6840
GACGTG	6846

(2) INFORMATION FOR SEQ ID NO: 199:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2911 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

GAATTCATTT TAAATAAAGA TACGGGAGAG GTAAGTGAAT TAAAACCTCA TAGGGTAACT	60
GTGACCATTC AAAATGGAAA AGAAATGAGT TCAACGATAG TGTCGGAAGA AGATTTTATT	120

1174

TTACCTGTTT ATAAGGGTGA ATTAGAAAAA GGATACCAAT TTGATGGTTG GGAAATTTCT	180
GGTTTCGAAG GTAAAAAAGA CGCTGGCTAT GTTATTAATC TATCAAAAGA TACCTTTATA	240
AAACCTGTAT TCAAGAAAAT AGAGGAGAAA AAGGAGGAAG AAAATAAACC TACTTTTGAT	300
GTATCGAAAA AGAAAGATAA CCCACAAGTA AACCATAGTC AATTAAATGA AAGTCACAGA	360
AAAGAGGATT TACAAAGAGA AGAGCATTCA CAAAAATCTG ATTCAACTAA GGATGTTACA	420
GCTACAGTTC TTGATAAAAA CAATATCAGT AGTAAATCAA CTACTAACAA TCCTAATAAG	480
TTGCCAAAAA CTGGAACAGC AAGCGGAGCC CAGACACTAT TAGCTGCCGG AATAATGTTT	540
ATAGTAGGAA TTTTCTTGG ATTGAAGAAA AAAATCAAG ATTAAGATAA AAGCTATAGA	600
AAAAAATGGT TTATGTACTG AGATTAGATA GTGAGGTGAT GACATAGTTT TGTGAAAATA	660
GCCATTTATA ACTCAATTAT TTAGTTTACT TTACTTTACT AGTGATACTA TTTGGAGTTA	720
TTAATGGACT TAGTTTATAT AACTAATGAA TTGATTGAAA GGGTTAGTAT TGACAATATT	780
GGTCATATTG ACTAGAAAAT AGAGTCTATC AAAATTTAAA GGCTAATAGA GGTGATGAGA	840
CAATTCGGC TCTTTGTCAA CTGTAGTGGG TTGAAGTCAG CTAAGCTCGA GAAAGGACAA	900
ATTTTGTCTT TTCTTTTMTG ATATTCAGAG CGATAAAAAT CCGTTMTTGG AAGTTTTCAA	960
AGTTTCGAAA ACCAAAGGCA TTGCGCTTGA TAAGTTTGAT GAGATTATTG GTCGCTTCCA	1020
GTTTGGCATT AGAATAGTGT AGTTGAAGGG CATTGACAAT CTTCTCTTTA TCTTTGAGGA	1080
AGGTTTTAGA GGATGAACTT GATTGAGATT GTCCTCAATG AGTCCGAAAA ATTTGTCAGG	1140
CTCCTTATTC TGAAAGTGAA AAAGCAAGAG TTGATAGAGA TTATAGTGGT GTTTCAGTC	1200
TTCTGAATAG CTCAAAAGTT TATCTATAGT AGATTGAAAC TAGAATAGTA CACCTCTGCT	1260
TCTAAAACAT TGTTAGAAAT CGATTTGACT GTCCTGAATG ATTTGTCCTG TTATTATTTT	1320
ATTTTACTAT AAATCCACGT TTACGAATCT CTTCCACAC TTGTTCAATG GGGTTCATCT	1380
CTGGTGTGTA TGGAGGAATA AATGCAAAAC CAATATTAGT CGGAATCTTT AAGGTACTTG	1440
ATTTATGCCA TATAGCATTG TCCATAACGA GTAAAAGATA ATCATCTGGA TAAGCTTGTG	1500
AAAGCTCCTA TTCCTAAAGC CCCTTTATAA CCTCTTGCGA GAGAGACTAT TGA CTCAGCC	1560
CTTACTTCAT GCGGATGAAA CTTCTTATCG GGTCTAGAG AGTCATAGCC ATCTGACCTA	1620
CTATTGGACC TTTTGTCTG GGAAAGTTGA GAATCAAGCA ATCAGCTGT ACCATCATGA	1680
TCAGAGTCGG AGTGGTTCGG TAGTACAAGA ATTCCTAGGA GATTATTCTG GCTATGTTCA	1740
TTGTGATATG TTGCGGCAGT AACTTAGGAC TTAGTCCTC TAGTTCTGCC TATGCGATAG	1800
CAGTCCAAGG TTAGGAGCA AGGCGACGCT AAGCTTGTA AACTGCGAAC CGCTAGAAGC	1860
TTATCGTCAA CTGGAAGAAG CTGAACTTGT TGGATGTTGG GCGCATGTGA GAAGGAAATT	1920

1175

TTTTGAAGCG ACCCCCAAGC AAGCAGATAA ATCATCCTTA GGAGCTAAAG GTTTAGCTTA	1980
TTGTGATCAG TTATTTTCCT TGGAAAKAGA CTGGGAGGCT TTGCCAGCTG ATGAACGACT	2040
ACAGAAACGT CAAGAACATC TCCAGCCCCT AATGGAAGAC TTCTTTGCTT GGTGCCGCCG	2100
TCAGTCAGTT TTAGCAGGTT CAAAAGTAGG AAGGGCAATT GAATACAGCC TCAAGTATGA	2160
AGAAACCTTT AAGACTATTT TGAAAGACGG ACATCTGGTC CTTTCCAATA ATCTAGCTGA	2220
ACGCGCCATT AAATCATTGG TTATGGGACG GAGTAAAGA GTCCAGTGA CTCTTTTAGC	2280
CTGAGCTCAG TTTAAAAAG CGAGGGTGGT TATTTTCTCA AAGTTTGAAG GGAGCTAAAG	2340
CAAGAGCTAT TGTTATGAGC TTGTTGGAAA CAGCTAAACG TCATCAATTA TAGTGCGTTG	2400
AATCTATAAC AGTACGCATC GACTGCTAAA ACATTTCTAT AAATCAATTT TCCTTTCCTA	2460
ATCGATTTGT TCATATCTTA TTTCAATCCA TTATAAATAG CGAGAAATAT CTATCCTATC	2520
TTCTAGAATG TCTTCCAAAC GAGGAACTC TCGTAAACAA AGAGGTTTGA GAGGTTTATT	2580
TACCATGGAC TAAAGTTGTA CAAGAAAAGT GCAAATAAGA AATCTCCAGA TTAGGAACATA	2640
TCCGTGAGTT CACTAATCTG GAGATTTTTC AATAGATTCG TTATTGGGCG GTTACGATAT	2700
GATCACTACT TCGTCAGTCT TATCTACAAC CTCAAAACAG TGTTTTGAGC AACCTGCGAC	2760
TAGCTTCCTA GTTTACTCTT TGATTTTCAT TGAATATTAG AACAGAAAAA ATGCTTGGAG	2820
TATTTGTTTG TGTGTTTATT TTTATATAAC AAATATAAA CAAAATAAAA ATATAAAAAA	2880
AGAGACAAAA AAGAACAGAA AGTAATTGAC A	2911

(2) INFORMATION FOR SEQ ID NO: 200:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6854 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

GAAAATAAGT CTTGACAGAA AGCGCTATCA ATGATAGAAT GAATTCAGAT AAAAAGATTT	60
ATTTTAAAA CAAAAATGAA ACGTTTCAAA AAAAGAAATA AAGAGACAGC GCCAAGCGCT	120
ATCTTTTCTA GAAAAAATG AAACGTTTCA AAAAGGAGG TTGCTATGAA TAGCAAAGCG	180
AAGCAAGTTT CTCTTTGGGA AAGAATCAAG AAACAAAAAC TCTTGTTATT GATGACTGTC	240
CCCGGTTTAG TTTTAACCTT TATCTTTAAA TACATCCCTA TGTATGGGGT TTTAATCGCA	300
TTTAAAGATT ACAATCCTTT AAAAGGAATT TTAGGGAGTG ATTGGATTGG TTTTCTGAG	360

1176

TTTACAAAAT TCATATCCTC TCCCAACTTT GGTATCTTGT TAGCCAACAC ATTAAAATTA	420
AGTATCTATG GTTTATTGCT TGGCTTTTFA CCACCAATCA TTCTCGCGAT TATGCTCAAT	480
CAACTCTTGA GTGAAAAAGT CAAAAACGA ATTCAGCTCA TTTTATACGC ACCAAACTTT	540
ATCTCAGTCG TTGTTATTGT CGGTATGATT TTCCTCTTCT TTTCAGTGGG AGGACCAATC	600
AACAATTTTC TTTCTATGTT TGAATGAAG GCTGACTTCT TGACAAATCC AGACTTCTTT	660
AGACCTTTAT ACATCTTTAG TGGTATCTGG CAAGGAATGG GCTGGGCTTC AACGCTCTAC	720
ACGGCAACAT TGGTAAATGT AGATCCAGCC TTAGTAGAAG CAGCCCGACT GGATGGAGCC	780
AATATCTTCC AACGAATCTG GCACATTGAT ATTCCAGCTC TTAAGCCTAT TATGGTTATC	840
CAATTTGTTT TAGCTGCAGG TGAATTATG AATGTCGGAT ATGAAAAAGC ATTCTTGATG	900
CAGACATCGT TAAATTTGCC AACTTCTGAA ATTATCTCGA CATATGTCTA TAAAGTTGGT	960
CTTGATCAG GAGACTATTC TTAACAACA GCGGTTGGTT TGTTTAATGC AGTGATTAAC	1020
GTAGTATTGC TTGTTGCAGT TAACCAATC GTTAAACGCA TGAATAATGG TGAAGGAATT	1080
TAAGGAGGAA AGTATGAAAA ATTCGATTAT GGATACAAAA TTTGATAGAC GTATCTTACT	1140
CTTAAATAAA ATCATTATTG TCTTTATCGT TTTGATGACT TTGCTTCCTT TACTTTTATAT	1200
CGTCGTAGCA TCCTTTATGG ATCCTAAGGT TCTGGTTAGT AGAGGGATTA GCTTTAATCC	1260
AGCCGATTGG ACTGTAGAAG GTTACCAGCG TGTATTCACT GACCAATCTA TTCTAAGAGG	1320
TTTTATCAAT TCTCTACTAT ACTCTTTTGG ATTTGCAGCT TTAACAGTCT TGCTATCTGT	1380
GTTTACAGCT TATCCTCTTT CTAAGAAAGA CTTGGTTGGA CGTCGTTGGA TTAACACTT	1440
CTTGATTGTA ACTATGTTCT TTGGTGGTGG TTTAGTCCCA ACTTACTTGC TCGTAAAAGA	1500
ATTGGGAATG CTCAATACTC CATGGGCTAT CATTTGTTCCA GGTGCTGTTA ACGTTTGGAA	1560
TATTATTCTT GCTAGGGCCT ATTTCCAAGG ATTCGCTGAA GAATTAGTTG AAGCTGCTGT	1620
CATTGATGGT GCAAATGATT TACAGATTTT CTTCAAAATC ATGCTTCCTC TTGCAAAACC	1680
AATTATGTTT GTTCTCTTCC TTTATGCTTT TGTAGGACAG TGGAACTCAT ACTTTGATGC	1740
AATGATTTAT ATCAAGGATC CAAACTTGA ACCATTGCAA CTTGTACTTC GTAAAAATCT	1800
CATTCAGAGC CAACCAGGTC AAGACATGAT TGGAGCACAA GCGGCTATGA ATGAAATGAA	1860
ACGTTTAGCT GAATTGATTA AATACGCAAC TATTGTCAAT TCCAGCTTGC CATTGATTGT	1920
TATGTATCCA TTCTTCCAAA AATACTTTGA TAAAGGAATT ATGGCTGGTT CACTTAAAGG	1980
ATAAAAAAAG AAAAAATAAA AGGAGTTTTT TCATGAAATT CAAAACATTC TCAAAATCAG	2040
CAGTTTTGTT GACAGCTAGT TTAGCAGTAC TTGCAGCCTG TGGCTCAAAA AATACAGCTT	2100
CAAGTCCAGA TTATAAGTTG GAAGGTGTAA CATTCCTCGT TCAAGAAAAG AAAACATTGA	2160



1177

AGTTTATGAC AGCCAGTTCA CCGTTATCTC CTAAAGACCC AAATGAAAAG TTAATTTTGC	2220
AACGTTTGA GAAGGAACT GCGGTCATA TTGACTGGAC CAACTACCAA TCCGACTTTG	2280
CAGAAAAACG TAACTTGGAT ATTTCTAGTG GTGATTTACC AGATGCTATC CACAACGACG	2340
GAGCTTCAGA TGTGGACTTG ATGAAGTGGG CTAAAAAAGG TGTATTATT CCAGTTGAAG	2400
ATTTGATTGA TAAATACATG CCAAATCTTA AGAAAATTTT GGATGAGAAA CCAGAGTACA	2460
<hr/>	
AGGCCTTGAT GACAGCACCT GATGGGCACA TTTACTCATT TCCATGGATT GAAGAGCTTG	2520
GAGATGGTAA AGAGTCTATT CACAGTGTC ACGATATGGC TTGGATTAAAC AAAGATTGGC	2580
TTAAGAACT TGGTCTTGAA ATGCCAAAA CTACTGATGA TTTGATTAAA GTCCTAGAAG	2640
CTTCAAAAA CGGGGATCCA AATGGAAATG GAGAGGCTGA TGAAATTCCA TTTTCATTTA	2700
TTAGTGGTAA CGGAAACGAA GATTTTAAAT TCCTATTTGC TGCATTGGT ATAGGGGATA	2760
ACGATGATCA TTTAGTAGTA GGAAATGATG GCAAAGTTGA CTTACACAGCA GATAACGATA	2820
ACTATAAAGA AGGTGTCAA TTTATCCGTC AATTGCAAGA AAAAGGCCTG ATTGATAAAG	2880
AAGCTTTCGA ACATGATTGG AATAGTTACA TTGCTAAAGG TCATGATCAG AAATTGGTG	2940
TTACTTTAC ATGGGATAAG AATAATGTGA CTGGAAGTAA CGAAAGTTAT GATGTTTAC	3000
CAGTACTGTC TGGACCAAGT GGTCAAAAAC ACGTAGCTCG TACAAACGGT ATGGGATTTG	3060
CACGTGACAA GATGGTTATT ACCAGTGTA ACAAAAACCT AGAATTGACA GCTAAATGGA	3120
TTGATGCACA ATACGCTCCA CTCCAATCTG TGCAAAATAA CTGGGGAAC TACGGAGATG	3180
ACAAACAACA AAACATCTTT GAATGGATC AAGCGTCAA TAGTCTAAAA CACTTACCAC	3240
TAAACGAAC TGCACCAGCA GAACTTCGTC AAAAGACTGA AGTAGGAGGA CCACTAGCTA	3300
TCCTAGATTC ATACTATGGT AAAGTAACAA CCATGCCTGA TGATGCCAAA TGGCGTTTGG	3360
ATCTTATCAA AGAATATTAT GTTCCTTACA TGAGCAATGT CAATAACTAT CCAAGAGTCT	3420
TTATGACACA GGAAGATTG GACAAGATTG CCCATATCGA AGCAGATATG AATGACTATA	3480
TCTACCGTAA ACGTGCTGAA TGGATTGTAA ATGGCAATAT TGATACTGAC TGGGATGATT	3540
ACAAGAAAGA ACTTGAAAA TACGGACTTT CTGATTACCT CGCTATTAAA CAAAAATACT	3600
ACGACCAATA CCAAGCAAAC AAAAAGTAGA GGTGATTAT GGGAGATAAG AAATACACAG	3660
TAGAAAAAGC CAATCGTTTT ATAGCAGAAA ATAAACATCT CGTTAATACT CAATATAAGC	3720
CTGAAGAACA TTTTTCAGCT GAGATTGGTT GGATCAATGA TCCAAATGGA TTTGTCTATT	3780
TTCTGGGAGA ATACCATCTC TTTTATCAAT TCTATCCATA TGATAGTGT TGGGGGCTA	3840
TGCACTGGGG ACATGCTAAA AGTAAGGACT TGGTGACTTG GGAGCACTTG CCAGTGGCAC	3900

1178

TTGCTCCTGA CCAAGATTAT GACCGAAATG GTTGTTCCTC AGGCTCTGCC ATTGTCAAGG	3960
ATGATCGCCT CTGGCTCATG TACACTGGAC ATATCGAAGA AGAAACCGGT GTCCGCCAAG	4020
TGCAAAATAT GGTATTTTCA GATGACGGGA TTCACCTTGA AAAGATTTC CAAAATCCAG	4080
TTGCAACTGG ATCAGACTTA CCAGATGAGT TGATTGCTGC TGATTTCCGT GATCCAAAAC	4140
TCTTTGAAAA AGATGGACGC TATTACTCCG TAGTAGCTGC CAAACACAAG GATAATGTGG	4200
GCTGTATCGT TCTACTAGGG TCCGATAACC TAGTAGAATG GCAGTTCGAA TCCATCTTTT	4260
TAAAAGGGGG AGAACACCAA GGTTTTATGT GGAATGCCC AGATTACTTC GAGTTAGATG	4320
GGAAAGATTG CCTTATTATG TCACCCATGC GTTATCAGCG TGAGGGAGAC TCATATCATA	4380
ACATCAACTC ATCGCTTTTG TTCACGGGTA AGGTAGATTG GAGAGAAAA CGTTTTATCC	4440
CAGAATCAGT TCAAGAAATT GATCATGGCC AAGACTTCTA TCGCCTCAA ACATTGTTGG	4500
ACGATCAAAA TCGTCGTATC CTGATTGCTT GGATGCAGAC ATGGGGCGGT ACCCTTCCAA	4560
CCCATGACCA AGAACACAAG TGGGCATGTG CCATGACTCT ACCTAGAATT CTAAGATTGG	4620
AAGATGGCAA ACTAAGACAA TTCCCTGTTA AAAAAGGCCA ATATCAAATC CAAATAGATA	4680
AAGATTGTCA TTACCACTTA GGAAATGATA TAGATTATCT TGAATTGGT TATGACAGTA	4740
ATCGCGAGCA AGTTTACATT GATCGTAGCC ATCTTATTCA AAAAATCTA GGTGAAGAAG	4800
AACAGGACAC TAGTCGACGG TATGTAGATA TTGAAGCTAA AGAATTGGAA GTTGTCTAG	4860
ATAAAAATTC CATCGAGATT TTTGTCAATC AAGGTGAAGC AAGCTTGA CTCAACTTATT	4920
ACTTAACGGT GCCAGCTGAG CTATCACGAA TTGATTAAAA ATTAAGTTAT TTCTCCTAAA	4980
GAAAAAGTTC TCTTTCTAAA ATAGTGGAAA GAGGACTTTT TGTGTTTGG GTATATAAGC	5040
TTAGTTTATG GTATTGTAA AATTGGTGTT GGATTATGAT TTAAGCTAGT TTTCTAAAGA	5100
ATTTGAAAAA AATTTTATTT AAGCAAAAAA ACCTTGCTTC CAAGGCTTTT CCTGTTGTAT	5160
TTAGATGCCC CCTACAGGGA TTGTAGGAGA TATGTTGCTT AGATGTTCTT GATTTTCTGG	5220
TGTTTGTAA CGTTTAAATG AGTTTGTGA GTTTGTGGT GGGGCGTTGC CCGGCAATTG	5280
CCCGACTTAT TGCTTGAAAA AGAATTAAAA ATATAGTATA GTTAATTATA GATTAACACT	5340
TGCTTGAGG AACTGATGAA GAACAATGAA AGATTAGGTA TTAAATTAAG TAGAGATAGC	5400
GTTTTAGGAT TGAGGGAAGT TAGAAGGCTT TATTTAGGCA GTTCAGATAT CCCAGTTTCT	5460
GATGGCTATG TGATTGAAGT TGCTTATAAC CAGATATCAC ATGAGATTGA TATTATTGAT	5520
TGGGTAGAGT TGAACAAGTC AAAAATTAAG ATAAGTGAAG TTAGTGAAAG CGTGGATATA	5580
GATGCCACTA GCTTGAGAAC AACTTGA CTAGACACAT TAGTATATGA AGGTATGAGA	5640
GATATACAGT TAAAGTTGAG AGAGCTTACA AAGGGGAGAG TATTCCTTTC ATTTGTAGTG	5700

1179

AAGTTAGTTT TGTTTGCTTC TATTTTAAAG AAAAAAGATT TACTAGAAAA ATTTCAAGAA	5760
AAGTGTTAAT CAAGTATTGA CACTTTATCT GGATTTCGGT ATAATATGCT TAGAAAGGAA	5820
TCTTTCTAAA TTTTTTCGT CCTTATGTGT TAATCAAAGA CGAATACAAA AACATATTTT	5880
TTTACTCTAA AAAGTGTTAA TCAATGATGT ATTTGTTAGA GAGGTAGATA AATGGAATTG	5940
AGAGCACCAC CAGTTATAAT AGTATAAAAC GTATAATAAA AATATTTTAA CTTGAATTAT	6000
AGAAAAGGAG AAACAAATCA TGAAACAAAA ACAACCGATT GTTTCTAGAA CGAAACAACA	6060
TACATTTGAA GAGCTTATTC AAGACCAAAA GTTAGAAAGA TTGGCTAAGT TGTCGCCCGA	6120
TTTGGTTGGA AGGTATGGTT TTAAGTCTAG CTGTGCGTCT TCATTGCGA ACTTGATTAA	6180
AGAAGCGTAT GGGGGTAAAA ATCTAAACGT AGTTTATGCG AGTCGGATGT TGGCTCTCTG	6240
GAATATTGCT TGCAGTTGTT ATCATAAGGC TGATGGGTAT TCTTTAGCAG ATGCGCTTTT	6300
TAGTGATAAA AAAATTTGTC TAGATTCTTA CTATTACCAC AAGAATACCT CTAATACCAT	6360
AACTAGTGAT GTGATAAAG ATGTTTACGA TAATTATAAT AATTATATGG TTTTAACTCG	6420
AGAAGCGACA CCTGAATACA TTTATGTTGT ACAAACCTGAA ATGCCAAAAG ATTCAGATTT	6480
ATATTTTAT ATTAGAGAAG TTCTGGGATT ATCGTTTAGT ACCATGCATT ATGCATTTT	6540
AGTCAAGGTT CTGTCAGGAG CGCTTGCTAG AAAATATAAG CCATATCGAA ATTGAATTAT	6600
TTAAATTTAT ACTCTTCGAA AATCAAATTC AAACCAAGTC AGCTTCGCCT TGCTGTAATC	6660
AAGTGCTGTC TGTGGCTAGC TTCTTAGTTT GCTTTTTGAT TTTCATTGAG TATTACTCTT	6720
ATGGTAGTTA TTTATGGCAT AATAATATTG ATTTGGGAGT TATAGCGAAA ATTTTAGGTT	6780
CTATAATATT TGTAGTGGGT AAACCACTAT AGATATTATG GAGCCTATTT ATTGTAGAAA	6840
AAAGTCCCAT ATGA	6854

## (2) INFORMATION FOR SEQ ID NO: 201:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3895 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

TCCTTGCTAA GTTTATACTC AATGAAAATC AAAGAACAAA CTAGGAAGCT AGCCACAGGT	60
TGCTCAAAGC ACCGCTTTGA GGTTCAGAT AAAACTGACA CGGTTTGAAG AGATTTTCGA	120
AGAGTATTAA TTTACATAAA TAGCCAGTGT TTGATAGGGT TTGAGTAGAA TTTTCTCAGA	180

1180

CACTTCTGCA TCTTCATAGT TTGATATCAA AATCTGTCCA TTTTGGTAGA CTGCTGGCAA	240
GTCGATTTCa CTTCTTTAGC ATAAAAGTTA TTGAGCACTA GTAACTTTTG ATCCTCAAAC	300
TGGCGTTCAA AAGCGTAGAC TTGTTTGCTA TCTTCAAAGG CTGGTTTGTA ACTTCCTTCT	360
GAAATGATTG GCATTCCTT ACGCATCGAA TCAAGTCTTG ATAGAAGGTA AAAATCGGAC	420
CCTGGATTTC ATTTTCTACA TTGATGTATT TATAGGATTT ACCAGCTTTC AACCAAGGAG	480
TGCCTGTTGA AAATCCTGCA TTTTCCGAAG CATCCCACTG CATGGGAATG CGTGAATTAT	540
CACGCGACTT AGCTTGAATA ATCTGGAAGG CTTCTTGCTG ACTCTTTCCT TCTCTAAGA	600
GCATCTGATA GGCATTAAGC GATTGACAT CCACATAATC AGCCATAGAA TCATAGTCTG	660
GGTCAATCAT CCCGATTTC TCACCCATGT AGATATAAGG TGTCCCACGT GACAGGTGAA	720
TGCTGGCTGC TAGCATGGTG GCTCCTTCCT TGCGGAAGTT TTGAATATCG ACAAACGGT	780
TCAAGGCACG TGGTTGATCG TGATTATTCC AAAAGAGGGC ACTCCAACCG TCTTTATCAC	840
TCATTTCTCT ACCCCAATA TGGTAAAGAC TCTTCAACTC TTCAAAATCA AAGGGAGCCA	900
AGGTCCACTT TTGTCCATCC TTATAGTCCA CCTTGAGGTG ATGAAAATTA AAGGTCATGG	960
ATAATTCCTG ACGATCAGGC GACGAATAGA GGACACAGTT TTCCATGGTG GTAGAAGACA	1020
TTTCCCCAAC TGTCAAAAG CTATCGTCGG ATCCAAAAGT GGCTTGGTTC ATCATACGCA	1080
AATAGTTATG AACGATGGGT TTGTCTGTAT AAGCTGGCTT CCCTTCATTT TCAGGACAGT	1140
CCACTGAAAC CTCGTCCTTA CCGATCAAAT TGATCACATC AAATCGGAAA CCTTTGACAC	1200
CCTTGTCGCG CCAGAAATTA ACAACCTTGA AAAGCTCCTT ACGGACATTG GAATTGCGCC	1260
AGTTAAGGTC AGCCTGGGTC TCATCAAATA GGTGAAGATA GTATTTCCCA GTATCCCCGA	1320
AAGGCGTCCA TGCAGAACCA CCAAACCTAG ACTGCCAATC TGTGGTTGG TCTTGGATGA	1380
AGAAAAAGTC TTGATAATAC TTATCACCAG CTAGGGCTTT CTGAAACCAT TCATGCTCTG	1440
TGGAACAATG ATTAAGTACC ATGTCCAGCA TAAAGTCAAT CTTGTGCTCT TTACCGACAC	1500
ACACCATTTT CTCAAAATCA GCCATATCAC CAAAAGAGG ATCCACTGCC ATATAATCTG	1560
AAATATCGTA ACCATTATCC CGTTGAGGGC TTGGATAGAA TGGATTGAGC CAGACCATAT	1620
CCACACCTAG TTTGGCTAAA TAGGGAATTT TTTGATAAT CCCACGAAA TCCCCAATAC	1680
CGTTTTCACT GGTGCTTTG TAAGATTTTG GATAGATTTG ATAGACTACT TTTCTTTAT	1740
CAAGTGTCAT CTGTTCTCC TTTTCTGATA AAAGGGAGGA AGCAGTCTTC CGTCCCTATT	1800
TGTGCTATTT CAATTATACT CAATGAAAAT CAAAGAACAA ACTAGGAAGC TAGCCACAGG	1860
TTGCTCAAAA CACTATTTTG AGGTTGCAGA TAGAGCTGAC GTGGTTTGAA GAGATTTTCG	1920
AAGAGTATTA GATTCGTGTA GCGACCATGA GAGATGCTCC AGCTTGGATC GTTGTGGAT	1980

1181

AAGTTCCGGG AATAGTCGCT GTATAAGCAT CTTGGTTGGT GATGATAACA GGAGTTTCTG	2040
TCACCAGACC TGCAGCCTTA ATGACATCCA TATCAAAACG AATCAGTTGC TGACCAACTG	2100
TAACGTGATC TCCTTGGACT ACAAGACTTT CAAAACCTTT GCCATCAAGA CCTACTGTAT	2160
CCATACCGAT GTGGATGAGC AATTCAACTC CCTCGTCAGA GACAATGCCG ATGGCATGCT	2220
TGGTAGGGAA AAGAACCGTC ACTGTCCCAT TAACTGGAGA GGTCAACTCA CCTTGGCTTG	2280
GTTC AATGAC TAGACCTTGC CCCATGACAC CTGATGCAAA AATAGGATCC GTCGCTTGAC	2340
TCAATTCTTT CACTTGGCCA GTTAGTGGGC TGATAATTC TACCGAAGTA AGTTCTACTG	2400
GTTTCATGGT CACAAATTCT GCTTCTTCTT GAGCAACGAA TTCTGCCTGC AAGTTCGTAT	2460
CGCCCTCTGT TTTTGTAAG AGACCAGCCT TGCGGAAGAA GAAAGTCAAG AGCATTGGAA	2520
CAACAATCGC AACTAGCATA GTTCCTGCAA ATGGCAGCAT GTATTGAGGT TGAATAGAGA	2580
GAATACCTGG CAAACCACCG ATACCAATAG AAGCCGAGT TACATTAAAA GTAACGGATA	2640
ACATGCCTGC AAGGGCTGAA CCAGTCATCC CAGCAACAAA TGGATAAATA TATTTTACGT	2700
TAACCCCAAA AAGAGCTGGT TCTGTAACAC CGAGATAGGC TGAAATGGTT GCAGGAAGTG	2760
AAACCTGAGC CTCACGCTCA TCATGGCGAT GCATGAAATA ATAGGCAAAC ACGGCTGAGC	2820
CTTGAGCAAT ATTAGAAAGA GCAATCATTG GCCATAGGGC AGTGCCACCA GCATCCGCAA	2880
TCAATTGTGT ATCAATGGCA TTGGTCATAT GGTGCAGACC TGTGATGACA AATGGAGCGT	2940
AGAGGGCGCC AAAAATTGCA CCGAAGAGCC ATTAACTGG ACCAGTTAAA CCTGCCAAGA	3000
CAACTGATGA AAGTCCTTGT CCAATTGTCC AACCATTGG TCCCAAAACA GTATGAGCCA	3060
AAATCAAGGC TGGAAATCAAT GACAAGAAAG GTACAAAAAT CATAGAAATG ACTTCTGGGA	3120
TATGCTTGTG CCAGAAGATT TCAAGATAAG ACAGACTCAA ACCTGCAAGC AAGGCTGGGA	3180
TAACTTGGGC TTGGTAACCG ATACGATTAA CAGTAAAATA GCCAAAATTC CAAACCCAGT	3240
TTGCCCGGAT ATCAGCTGCT GGCCTTGAAG CAACCGCATA GGCATTGAGC AACTGAGGTG	3300
ATACCAAACA GATTCCGAGA ACAATTCCA AAATTTGGCT GGTTCCTATC TTACGAGAAA	3360
CAGACCAAGT AATCCCTACT GGTAAGAACT GGAAGATAGC TTCACCAGGC AACCAGAGGA	3420
AGTGATTGAC ACCTGCCCAA AACTGAGAGG ATTCTGTGAT GGTCTTGCCA TCCAACATCG	3480
ACCAATGGAC ACCTTCCAAG ACATTACGGA AACCAGAGAT CAATCCTCCG ACTATCAAGG	3540
CTGGAATAAT CGGAGTAAAA ATCTCCGCCA GAGTGGTCAT AACACCTGG ACCACGTTTT	3600
GATTACTCTT AGCTGCAGAC TTGGCTGCTT CTTTGGAAAC ACCCTCAATA CCTGAAACGG	3660
CTGTAAATC ATTATAAAG ATGGGCACGT CATTTCCAAT GATTACCTGA AATTGACCTG	3720

1182

CATTTGTA	AA	GGTTCCTTTA	ACAGCTGGAA	TTGACTCGAT	AGCTTTAACA	TTAGCCTTCT	3780
TATCATCTCC	TAAACAAAC	CGCATCCGTG	TCGCACAGTG	AGTTACGGCA	GTCACATTTT		3840
CTTTGCCTCC	GATTGCCTGA	AGCAGATCTT	TGGCTTCTTG	TTCAAATTTT	CCCCG		3895

## (2) INFORMATION FOR SEQ ID NO: 202:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3936 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

AGGATCGCCG	CTCCAGCTAC	TAAGTCTCGT	GCAGTGCCGA	TTTATCAAAC	AACATTTTTT	60
GTTTTTGATG	ACACGTAGGA	AGGTGCCGAT	CTGTTTGCCT	TGAGGAAACC	AGGGAACATT	120
TATACTCGTA	TCACCAATCC	TACAACAGCT	CCCCTTGAAG	GTGGTGTTGA	AGCGCTAgca	180
ACAGCATCAG	GTATGACTGC	AGTGACTTAT	ACGATTTTGG	CGATTGCCCA	TGCTGGTGAC	240
CATGTAGTGG	CTGCTTCGAC	TATTTACGGT	GGAACCTTCA	ATCTTTTGAA	AGAACCCCTT	300
CCTCGTTATG	GTATCACAAC	AACCTTTTTT	GATATTGATA	ATTGGAGGA	AGTAGAAGCA	360
GCTATCAAAG	ACAATACCAA	GCTTGTCTTG	ATTGAAACCT	TGGGTAACCC	CTTGATTAAT	420
ATTCCAGACC	TGGA AAAACT	GGCAGAGATT	GCTCATAAAC	ATCAAATCCC	ACTTGTGTCA	480
GACAATACTT	TTGCAACACC	TTATTTGATT	AACGTCTTCT	CTCATGGCGT	TGACATTGCC	540
ATTCACTCTG	TGACTAAGTT	TATCGGTGGG	CATGGTACAA	CTATTGGAGG	AATAATTGTC	600
GATAGTGGTC	GTTTGTACTG	GACGGCTTCA	GGGAAATTCC	CTCAATTTGT	TGACGAGGGT	660
CCAAGCTGCC	ACAATTTGAG	CTATACTCGT	GATGTGGGTG	CAGCAGCCTT	TATTATAGCT	720
GTTTCGAGTC	AATTGCTTCG	TGATACAGGT	GCAGCCTTGT	CACCATTCAA	TGCTTTCCTC	780
TTGCTACAAA	GACTTGAAAC	CTCTTCACTT	CGTGTGGAAC	GCCATGTACA	AAATGCTGAG	840
ACAATTGTTG	ATTTTCTTGT	CAACCATCCT	AAGGTAGAAA	AGGTAAATTA	TCCAAAACCTT	900
GCAGATAGTC	CTTATCATGC	CTTGGCTGAG	AAATACTTGC	CAAAAGGTGT	CGGTTCAATC	960
TTTACCTTCC	ACGTCAAAGG	TGGCGAGGAA	GAAGCACGCA	AGGTCATTGA	TAATTTAGAA	1020
ATCTTTTCTG	ACCTTGCAAA	CGCGGCAGAT	GCTAAATCGC	TTGTTGTCCA	TCCAGCAACA	1080
ACCACTCAGC	GTCAATTGTC	AGAAAAAGAC	CTAGAAGCAG	CAGGTGTCAC	ACCAAACATA	1140
ATTCGTTTGT	CAATCGGTCT	TGAAAATGTA	GAAGATTTGA	TTGAAGACTT	GCGCTTGGCC	1200
TTGGAAAAAA	TTTAAAGTAA	AAGAAGATAA	ACAGTGGGCT	TCGACTCACT	GTTTTTGATT	1260

1183

TTCCCTCAGG CATGATATAA TGGTTACAGA AGTCTAGAAA GAGGAACGAT ATGAACGAAA	1320
TCAAATGTCC CAACTGTGGG GAAGTCTTTA CAGTAAATGA GAGTCAGTAT GCCGAACTCT	1380
TGTCCCAAGT GAGAACGGCA GAGTTTGATA AGGAACTACA CGATAGGATG AAGCAGGAAC	1440
TGGCCTTGGC TGAGCAAAAG GCCATGAATG AGCAACAGAC TAAACTGGCT CAGAAGGATC	1500
AAGAAATTGC GCAATTACAG AGTCAGATCC AAACTTTGA TACAGAAAAA GAATTGGCCA	1560
AGAAAGAGGT TGAACAGACA AGCCATGAGG CTCTCTTGGC TAAGGACAAG GAAGTACAGC	1620
TCTTAGAAAA TCAGTTGGCT ACCTTGCGTT TGGAGCATGA AAATCAACTA CAAAAGACCC	1680
TTTCTGACCT AGAAAAAGAA CGGGATCAGG TTA AAAACCA ACTACTTTTG CAGGAAAAGG	1740
AAAATGAATT ATCTTTGGCT TCTGTTAAGC AAACTACGA AGCCCAGCTC AAGGCAGCTA	1800
GTGAACAAGT CGAGTTTAT AAGAATTTA AGGCTCAACA ATCTACAAA GCGATTGGGG	1860
AAAGCCTAGA ACAGTATGCA GAGAGTGACT TTAACAAGGT TCGTAGTTTC GCCTTTCCAA	1920
ATGCTTACTT TGAGAAGGAT AACAAGGTCT CTTCCGCTGG GTCTAAAGGG GACTTTATCT	1980
TCCGTGAGTG TGATGAAAAT GGAGTTGAAA TCATTCTAT CATGTTTGAG ATGAAAAACG	2040
AAGCGGACGG AACAGAGAAG AAGCACAAGA ATGCAGATTT TTACAAGGAA TTGGACAAGG	2100
ACCGTCGGGA GAAGAACTGT GAGTATGCCG TTTTGGTGAC CATGCTTGAG GCTGATAATG	2160
ACTACTTTAA CACAGGGATT GTTGACGTCA GTCACGAGTA TGAAAAATG TATGTTGTTC	2220
GTCCTCAATT CTTTATCCAA TTGATTGGTC TCTTACGTAA TGCGGCGCTA AATTCCCTAA	2280
AATACAAGCA GGAGTTGGCC TTGGTTCGCG AGCAAAATAT TGACATTACG CATTTTGAGG	2340
AAGATTGGA TGCCTTTAAG CTAGCTTTTG CTAAGAACTA TAATTCAGCT TCGACTAACT	2400
TTGGAAGGC TATTGATGAA ATCGACAAGG CCATCAAACG CATGGAAGAG GTTAAGAAAT	2460
TCCTGACCAC ATCTGAAAAC CAACTCCGTT TAGCTAACAA CAAATTGGAA GATGTCTCTG	2520
TTAAAAAATT GACCCGGA AATCCAACAA TGAAAGCGAA GTTCGAAGCA CTGAAGGGGG	2580
AGTAGAAAGC AAAAATGAAC GGTATTATTA ACTTAAAAA GGAAGCAGGA ATGACCTCGC	2640
ATGATGCGGT TTTTAACTG CGTAAGATTT TGGGAACCAA GAAAATTGGT CATGGTGGA	2700
CCTTGATCC GGATGTGGTG GGTGTTTTGC CGATTGCGGT TGGCAAGCG ACACGCATGG	2760
TCGAGTTTAT GCAGGACGAG GGTAAAGATCT ATGAGGGGGA AATCACTCTG GGCTATTCCA	2820
CGAAGACTGA GGATGCTAGT GGGGAAGTGG TCGCAGAAAC CCCTGTTTTG TCTCTCTGG	2880
ATGAAAAGCT TGTGATGAA GCGATTGCTA GCTTGACTGG GCCTATTACT CAGATTCGCC	2940
CTATGTATTC GGCAGTTAAG GTTAATGGTC GCAAGCTCTA TGAGTATGCG CGTGCTGGTC	3000

1184

```

AGGAAGTGGG GCGTCCAGAA CGTCAGGTGA CCATTATATCA ATTTGAGCGA ACAAGTCCGA 3060
TTTCTTATGA TGGCCAACTT GCCCGATTCA CTTTTCGTGT AAAATGCAGT AAAGGGACGT 3120
ACATCCGTAC TTTGTCAGTT GATTTGGGTG AAAAGCTTGG TTATGCGGCT CATATGTCCC 3180
ATTTGACTCG TACTAGTGCT GCTGGCTTAC AATTAGAAGA CGCTCTTGCC TTGGAGGAAA 3240
TTGCTGAAAA AGTAGAGGCT GGGCAATTAG ATTTTCTCCA TCCTTTAGAG ATTGGGACAG 3300
GTGACCTTGT CAAAGTTTTC CTAAGTCCAG AAGAGGCTAC AGAAGTTCGC TTTGGTCGTT 3360
TTATTGAGCT AGACCAAACG GACAAAGAAC TGGCTGCCTT TGAAGATGAT AAATTGTTAG 3420
CCATTCTAGA AAAACGGGGC AATCTCTATA AGCCAAGGAA GGTTTTTAGC TAGATCGTTT 3480
AGGAATAAAA ATCGGGTGAT AGATAACAAT TGCTTGATAA AACCCCATAC TAATAGTAGA 3540
ATGGTTTTGG GAATTATAAT ATTCCAATTG TTGCGAGTTG TAGGTACTCA AATAATCTAT 3600
ATAGAAATTT AGAGGTGTGA AATGAAGCAA TTTAAAATTC TTTGAGATAA ATATTTAGAG 3660
TCCATTACAG GTTCTGATGG GAACTTAGGC CCAGGATTTG GTGTGATAAT TCCATGATGC 3720
GAAATGAGTT TCGAGAAAGG GTGGAGCAAC TTCTTCAACA AAAAGAAATA AATGAAAATA 3780
GTGAGTTGAG TCACCTGTTT CGTCTTGCTA TACAAAATTT AGACAGAAAT GAAAAATACC 3840
AATCGGTCAT GGCCAATTTG AGTCAAGGGT TGTCACTTTA CCTCATGACG CATCATTACC 3900
AGGCACCTAA GTCTGTCATT GATTTTGGTT TATGGA 3936

```

(2) INFORMATION FOR SEQ ID NO: 203:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3230 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

```

CATCCAGCAA CTGCTCCTCT GAGCGTTTCA AAATGATGT AATTTTCTA GTTTTTCTA 60
ATAAATGTGC CATTTTTCAC CTCGAATTTA ATCGCTATCA TTATAACATA AAAACGCTC 120
TTTTCAATA ATTATCTGAA AATTCCTTAT TGACTTGCAT TGACTTACAA TTTAATTAAA 180
AACCAGAATA TTTTAAATTA AATTGTTCCT TTTCTATTGA CAAGTGCCT ATTTTGTGT 240
ATCATAATAT TATAAAGAT AATATAATAA TTTTATTGT CTTTTCACAT TCGGCTCCT 300
TATATAAAAA AGCGATTCAT TTTGAACCGC TTTTCTTAT TTATCGCCTT TGTTACGAAT 360
AACAAAGCCT GTTTGCTTTT CGCTTAAAGT ATTGCGTGGT TTTTATTAT CCTTACGGTA 420
ACGTTTTTCC TTATCAAAAC GATCGTTGCC ACGACTTCCT TTTTGAAC TATCAGGCG 480

```



1185

ACCATTGCCA CGGCGATCAC GCTCTCGACG GTCGTCCCCA CGACGGCCTC CACGACCTCC	540
CTTAGCTTTA CCACCGAAAC CATTACCTGA TGGTTTAAAC GGTAGTGGGT TTTCACGTGC	600
AATCTCCACT TCTGGAAGGC TATCTGGGTC TTGGACTGTC AGACTCAAGA TATACATTGC	660
CAATTCTTCT GGAGTAACT CAGCAGCCAA TTTGCGAGCA TCCTTACCAA ATTTCTCAAA	720
GTTGGCACGA ATGGTTTCAT CTGCAAAATC ACGTTCGATT TTCTTGAGAG CTACCTGTTT	780
TTTGTATTGG AAGGATTCTT CTACACTTGC AGGTTTGAGA CCTTTCATGC GTTCTTAGT	840
CAAGTTTTC AATGATTGAA GGTAAACCAT TTCGTTTGA GCAACAAAAG TAATAGATTG	900
ACCTGACTTA CCAGCACGAC CTGTACGACC GATACGGTGA ACATAACTCT CAGGATCTTG	960
TGGAATATCG TAGTTGTAGA CATGGGTAC ACCTGAAATA TCCAAACCAC GCGCTGCAAC	1020
GTCTGTGCGA ACCAAAACAT CAAGATTGCC ATTTTAAAG TCACGAAGGA CACGAAGACG	1080
TTTGTMTTGG TCTAGGTGCG CATGAATTCC TTCTGCACGG AAGCCACGAA TTTTCAAACC	1140
ACGAGTCAAT TCATCCACAC GCGTTTGGT ACGACCAAAT ACAATAGCGA GTTCTGGTTG	1200
TGCCACATCC ATGAGACGAG TCATGGTGTC AAATTTTCT TGTTCCTTAA CACGGATATA	1260
GTA CTGGTCA ACCAATTCTG TTGTCAATTC CTTAGCCGCA ATCTTGACAT GTTCAGGGGC	1320
TTTCATAAAC TGAACACCGA TACGTTTGAT GGCATCTGGC ATAGTTGCTG AGAAAAGCAA	1380
AGTTTGACGG TTCTCAGGTA CACGGGAAAT AATGGCTTCG ATGTCTTCAA GGAAGCCCAT	1440
GTTAAGCATT TCATCCGCTT CGTCAAGGAT AAGGGTTTCA ATGTCTTGTA ATTTCAAGGC	1500
CTTGCGTTTA ATCAAGTCCA AGAGGCGACC TGGAGTTCCC ACCACAATAT GGGCACCAGA	1560
TTTAAGAGCC TTAATTTGTT TTTCAATGCT TGATCCGCCA TATACTGAAC GGACTTTGAC	1620
TCCCTTACTA CGACCAAAGC GGAAGAGTTC TTCTTGACTT TGGACAGCTA GTTCACGAGT	1680
TGGAGCGATG ACCAAGGCTT GGATAGTCGC TTCTTCTGTA CGGATTTTTT CAAGGGTAGG	1740
CAAGCCAAAG GCTGCAGTTT TTCCTGTACC AGTCTGAGCT TGACCGATAA CATCCTTGCC	1800
TTCAAGGGCC AAAGGAATAG TTTGTTCTTG GATAGGACTA GCTTCTACAA AACCAGCTTT	1860
TTCAATTTCT GCTAGCAAAT CAGCAGACAA GTTTAATTCA TTAAATTTC A GTTATTCTT	1920
CTTTCTAAAG GTGGTGCGAA GCCACCCTAT AGGGCTTAGT TTATACTTTT CTTTTATGA	1980
CGTATTTTCA TATACTAGA TATAAAATCG TGTGCTTCT TTTCCACAAA AGAAAAGTAC	2040
TGTTTTCTTT GCAACCTATC TAGTATAACA CAAGACCAGA GCAAAAGATA GCCCATTTT	2100
TACAGAAAAT CATGTAAGCG CTTTTTGACT TTCTTTTTTG ATTGAACGAC CTAGATAATA	2160
AGACAAAGCC AAGGCGATAC TGTATAAAAT GAGAAAAACG AACAAGGTTT GTGTGTACGA	2220

1186

ATGAGCCATT TTATAAGTCT CTGCTAATAA AATAGGTCCC GCTAAACCAG CCATTGCCCA	2280
AGCTGTAAAA ATATAACCAT GCAGAGCGGC CAATTCCTTG GTTCCAAAAA TATCACTGAG	2340
ATAAGCTGGA ATCAAAGAAA AACCAGCTCC ATAGCAAGTC ATCAAATAG ACATAGCAAC	2400
TACAAATAAA ACGGAATCTG TAAAGAGCCA AAGTGAGAGA GAAAAGAAAA GATTGACAAG	2460
CAGTAATATA CTAAAGGTTA GAGGGCGACC GATATAGTCA GACAACTCG CCCAGAGCAA	2520
GCGACCAAT CCATTGAAAA TCCCCAAAAC ACCCACCATT ACTGCTGCAT GACTTGTAGA	2580
CAAGCCAGCC ATCTCCTGTG CCATTGGCGA TGCCGCTGAA ATTAAGCCTA AACCACAAGC	2640
TATGTTGATA AAGAAAATAA TCCAAAGCAT ATAAAACCGA TTGCTTTTPTA GAGCCTGATT	2700
TGCAGCCATT CCTTGCGTCA AAGAGGCTGT TTTTCTTTC CCTGAAGAAG ATAAAAATTGC	2760
AAGCTCTTGC TCATTTGGAC GCTTAATGAA TTGTGAAGCT AGGAGCATGA TAATAAAGTA	2820
ACTTGCTCCT AAAATATAAA AAGTTTCTAC AAGCCCTACC CCTGCGATGA GGTGTTGCGC	2880
TATGGGACTA GTCAATAAAG AAGCAAAACC AAACCCATA ATCGCTAAAC CTGTTGCGAG	2940
ACCACGTTTA TCAGGAAACC ATTTTATAAT CGTCGACACA GGGGTAATAT AGCCTGCTCC	3000
CAAACCAAGC CCACCTAAAA TGCCATAAGC GAGATACAAC AACCACAGCT CTGACGGTCT	3060
ATTGCAAATC CTGTTAAGAT ATTTCCACCT GCGTATAGAA AAGCAGATAG ACTTCCCATG	3120
ACTTTCGGAC CAAATTTTTC TACCAAACGC CCCATAAATG CAGCCGATAA GCCCAAACAA	3180
AAGATTGCTA GACTAAAGGC GAAGGCAACA GAAGCCTGAT CCCATCCCGT	3230

## (2) INFORMATION FOR SEQ ID NO: 204:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5096 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

CCTATGAAGA CTGTCCCAAC TGGGTGTCCT TCTAGGCTAT CTGGTCCTGC CACTCCAGTC	60
AAACTAATTC CAAAATCAGA CTGGGTCTTG CTTCGTGCCT GCTCAGCCAT CTTCTGAGCT	120
GTAAATTCAG ACACCACACC ATGTTCTTCC AAATTCTTGG CAGGAATATC CAACATCCTT	180
GATTTTTCCT CCAAGCTATA GGTCACAAAA CCACCCTTAA ATATACTTGA AACTCCAGAA	240
AAATTCGCCA CGGTAGCTTG GAAAAGACCT GCCGTCAAAC TCTCTGCAGC CGCGATGGTT	300
TTCCCTTGCC TTTTCAGTTC TTCTACCACA ATGCTGGCTA AACTAGTTTC TTCCCCATAA	360
CCATAGCAAA AGTCTCGTAA AGAAATTCCT TCGAAAGTCT GGCAGTCCAA GATTGATTT	420

1187

TCCAAGATAT CCAGCGCTTG ATTCGCCTCT TCTTGACTGC TAGCCTTTGT TGACAGACGT	480
AGAGTGACTT CTCCTGTCTT GGCATAAGGG GCCAAGGTAG GATCGATCTG ATTATCAATT	540
AAATCAGCCA AAATCGTAAC CAACTGGCTC TCGCCAATCC CAAAGAAACG AAGAACTCGG	600
GAATACAGCT TGCTCCCTGT CATCAACTTG GGTAGAAGTT GGTTTAAGAC CATGGGTTTC	660
AATTCACTTG GCGGACCTGG AAGGACGACA TAGGTCACTC CGTCTACTTC TAATTTTCCT	720
CCAACAGCCA GTCCTGTTTC GTTTGGCAGT GGAATCGCTC CTTCTACAAT TTGAGCTTGT	780
CTTTCGTTAT TCGGTGTTTC GGCATAGTCT GGTGCGAGGG TAAAAAAGAT ATCCAACTTC	840
TCCTGAGCCT GAGGATCAAA GACTAATGCT TTCCCTAAAA ATTTAGCTAG GGTTCGTTTG	900
GTTAGGTCGT CCTCAGTTGG CCCCAAACCG CCTGTCAAAA TCACCAGACT GCTACGTTGA	960
CTGGCAATCT CAAGCAAAGA CAAGAGACGA ACTTCATTGT CTCCTACAGC CGTCTGAAAA	1020
TATACATCTA CCCCATCTC AGCTAGTTT TCCGACAAAA ACTGGGCATT GGTGTTGACA	1080
ATCTGCCCTG TCAAAATCTC TGTTCCAACA GCAATGATTT CTGCTTTCAT GTTTCCTCCT	1140
ACCTATCTAT TCGTATTTTT TTGAAAAAT CGCAGGAATT TTCCTACGAT TGATTTTTTT	1200
ATTTGTATCA AAAGTTAATT ATCTTCATCA CCAACAGGTG CTCTGCCAAA TAAATCTTCA	1260
AATAAAACCG CATTGGTTTC AAGCTGAGTA ACTTCTTCTT GTCCCAAAGA ACGTCGGAGT	1320
AGATTTTGCA TTTCCAACAT ATGTGCTCTC GAAACAATCT GGTAAAGAAC ACCTGAAGT	1380
ATCTCTCCTT CACCCTGCAA CTGCTGAGTT TCAATGGTTT TAAATGAATC TTTATAGCCT	1440
AGCAAGTTAG GGATACTTTT TGCAGACAAA TCAATATTGG TCTGCATATT GTCACTCAAA	1500
GCTTTTAGAA TCTCTTGATA ATGACCAATG CTATTTAAAC TGAGAGCTTT TTCCATGACT	1560
TTTTGAATAA CTTACAGTTG ACGTTTTTGA CGACCATAAT CCCCCTCAGG ATCTGGTAA	1620
CGCATTCGTG CATAGACTAG GGCTTCTTCT CCCCCAATAT GTTGCTCCCC AACACCGATA	1680
GAAATAGTAT TAAATTCTTC TTGGTCACTG ATAGAAATTG GGAAACCTAG GATATTATTG	1740
ACTGTAATAC CTCCTACTGC ATCCACTAGT TTTTGCAATC CTCTCATATT GACCATCACA	1800
TAGCGATCAA TATGGATATT CATCATTTTT TGAATGGTTT CTATAGCAAG CTCTGCTCCA	1860
CCATCTGCAT ATGCTGAGTT CAGTTTCGCT TCATGAGCCT GACCATTCCC TGATTCAATG	1920
CGCGTCAGAA TATCCCGCTC TAAACTCATC ATTGTTGTTT TTTTCGTTTT AGGATTCACT	1980
GTCATCAAGA TCATGCTATC ACTTCTACCG ACCCAAGTTT CAGTTCGTTT AACATTTCGG	2040
GTGTCCACTC CCATTAACAG AATGGTTAGA GGTTCAGTCG CTTCAATAAC CTTGGTTTCT	2100
TCACCGATTT TTTTATAGGT TTTAGCTAAG GTTCTGTCC CTTGTTGATA AATAGTATAA	2160

1188

GCAAAAACAC CTACTCCTAC TACAGTTACA GAAAGTAAAG CTAGCACCAT TCCAATAATT	2220
TTTTTAACCA TATTTCTACT AACCTATCAG TTTACCCATC AAGTAAACAT CGATAAATTT	2280
CCCTTCTTCT ATATATGCCC CACGCTCTTG GCTACCTTCA ATGACAAAGC CATGCTTTTG	2340
ATAAAGATGG ACTGCTGCTT GATTACGAGT TTGGACAGTC AGTTGGAGAC GACGCAGAAT	2400
GCCACTTGCT TGTGCCCACT CTATCGCTTC TTCTAGCAAC AAACCTCCCA AGCCATTATT	2460
CCAATATCTT TTCCAATCA CAATGAAGAG ATCTCCAATA TGACGGACTC TCTTACGCTG	2520
ATCAGCTGTA ATATTTACAA TACCAGCAAT TTTGCCATTT AAGAATGCAA GTAAGGTTAT	2580
CTGATTGTCC GAACTAGCTT GCTTGTGAG GAATATTTCC ATCTCCTCAC TAGTCAAGAG	2640
AATACCATCT CCGTCTAGGC TGGTAAAGTC TGTCTCCAAA CTCACACGAT TAAAAAAGGC	2700
CACTAATTCA GCTGCATCTT TGGGCTCTGC TTCCCTAATG AGCAATTCAT ACTCCATATT	2760
GAAGCTCCTC TAACAATTTC TCAGCACGCA AACCCTTTGC CTGAAAATTT AAACGGCGTC	2820
CATCTGCTTC TTTTAGAATT TCCAATTCTA AATAAGCATC TGGCAAGGCA TCTCCTAAGA	2880
GATTTCCCA CTCAATAACA GTCACGCCGC CACCAAGAT AAACCTCATCC AAGTCGATAG	2940
AATCAGCATC TCCTTCAATA CGATAAACAT CTAGGTGATA AAGTGGAAGT CGACCTTCAT	3000
ACTCTCTCAC GATAGTATAG GTGGGACTTT TAATCATTTG AGAAATCTGT AATCCTTTTG	3060
CAAGTCCTTT AGTAAAGGTC GTTTTACCTG CACCCAGTTC TCCAGTTAAG ATTAACAT	3120
CATTCTTTGC TAATAGATGG CCCAACGCT CCCCTAAGGC TTGCAACTCT TCTTCATTTT	3180
TTGTGTACAT ACTCTTATTA TACCAAAAAC TTTTCTTTTG TGTCTATTTT CCTACTAAAC	3240
TTATCATCAT AACATCCATA AAAAACAGGC TTTCTCTAAA AGAAAATGAG CGTAACAATG	3300
ACCAATACAA GATCTCGGAA AATATGACCA TAAAAGGAAA CTTCCTTCTT AACCGAATTT	3360
GGGACAAGAT AGGCTGCAAA AAACAAGCCC AGTCCAATAT AAATCAGAAG TGAGACAATG	3420
GTCATTGGAT TTCTTAAGAA AAGAAGTGT GCTAAAATAG TCACCAACAC TGTCTTTTTT	3480
CTGTCCAGCA TAGCAAGAAA ATCGCGCAG TATTTTTTCA AGGGTAAAAA AATCAGCAAA	3540
TCTAGCCCAA ATAGGAAAAA GAAGGATGGC AATAAAAAGT CAACTAATTC TTGCTGCAGC	3600
GTATTTTTGA TGAACAAGTT ATCTGACAAA ACAAGAACAG CTCCTAACAA ATTAATTAAG	3660
AGTAACATAC TGTA AAAAAG CTTCACCGAC TTCTTACTGG CTAGGACACT ATGGACTTCT	3720
TGCTTACGGG TATAAAGATA ATTTACTCCA GCACAGATTC CTGAAACGAA AACCATGCTT	3780
CCGATGAAAA AAGCTGTACT TTGTTTAAAG GACAAGATGC ATTCTTCCA TAGGAAACAG	3840
CTACTCAAAC TGATTTGAAT TAAAGCTAAC AAAAATAAGA TTCTCATGA TTTTCATCTC	3900
TCTCTCCCTT CCTACCAATC ATTATACTAG GAGAAAAGAG AGAACTGTTT CTAATCTTCT	3960

1189

CAAATGTCTC TTTAAGACGC TAAACAAACA CTAGAGACTA ATACTCAATG AAAATCAAAG	4020
ATCAAAC TAG GTAGCTAGCC ACAGGTTGCT CAAAACAGTG TTTTGAGATT GCAGATAGAG	4080
CTGACGTGAT TTGAAGAGAT TTTCGAAGAA TATAAATTTG AAATCATGAA AATCCGTCAA	4140
ACGGGTGGTT GTTTTGTCTC GCACCTCAGC GAGCGAGACG GACTCAGACT CACATAATTA	4200
TAAGGCTGAT AGTATTAATC TAACTATCAG CcTmCAGGTT ATTTAACGTT TCAGAAAAAC	4260
TATAATGTCA AGATTAAC TAACAGTATCT AGTTCCTTCA AATAATTTTC TATCTTCATC	4320
AACATTAAAG GATTGTTATA AATCTTACAT AACTCTCTTG CTTCTATATA ATAATTTTGT	4380
ACTTGTTCTC TGTCTAGAAA TTTGGCTCCA GCATTTCTTA CAAGAATAAG TAGAGGAGCC	4440
AATTGGTAGC TTGTCTGTCT TTGTTTACAG AGTTCAATCG TTTCAAGAGC TTCTTGGATG	4500
GCTTCATTAT ATTTTTCCTT TGATACTAGG TAGTGAGCGT AGTTGTAACG AACTCTGATG	4560
TAGCCAAATA AAAACTCTTG ATGGTCCAAA TTTTGTCT GATACAACTC TATTAAATGA	4620
GAGTAGTTTG CCTCATATTC TTGTTACGA CCCACTAAGG AATAGAAATT AGATAGAGTA	4680
TTCAACGCCT TTAAATAAAT CAGAGTATTT GAAGAGACTT TTAATAATAT ATTTTCCAAT	4740
GACGAAATTG CCTCACACTT ACTGTCATAT TGATAGAAGT CAATTATAGA TTTAATCCAT	4800
TCAAGGTAAG TTCGGTCTTC TAATGTTAGA AAAGTGCTTC GTTCTACTTC TATTTTATAA	4860
AGATATTCTA AATCGTCATA ATTTCTGTCA TCTAATAGGC GAGCAGATAG ATGTTTGAAA	4920
TTAGAGAGGT TAGACTTAAC TTCGATTGT TCATTGAAAA AGTAATCCAA AGGGACTTCA	4980
AGTCGTTGAG AGAGTTTGAA TAACAAGTCT GCGGAGGGAA TAAATGACC TCTTTCAATT	5040
TTACTAATCT GGCTTTGTTT ACAAATTCCT TCTGCAAGAG TTTGTTGGGA GAGTCT	5096

## (2) INFORMATION FOR SEQ ID NO: 205:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2395 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

ACAAGATAAA AATAAAGGAT TACAATGGGG AATATAAAGT AAACCGGTAA ACCTAAAAAG	60
AAAGGAGAAA AGATGAAAAT TGTACTTGTA GGCATGGAC ATTTTGCTAC AGGGATTAT	120
AGTTCTTTAC AATTGATTGC AGGTAATCAA GAAAATGTGG AGGCGATTGA CTTTGTGGAA	180
GGAATGTCAG CAGATGAACT CAAGCAAAAA ATCTTACTTG CAATTTCAAA TGAAGAAGAA	240

1190

GTTTAAATCC TAAGTGATCT CTTGGGAGGA TCGCCATTCA AGGTTTCTTC TACCATAATG	300
GGAGAAAATC CAGCCAAGAC AATGAATGTT CTCTCGGGTT TGAACCTAGC CATGTTAATG	360
GAAGCAGTCT TTGCTAGAAT GGCTCATAGC TTTGATGAGG TTGTTAATAA ATCAGTAGTG	420
GCGGCCCAGG GCGGAGTCGT AAATGGTAAA GAATTGTTTT CAACGGATGC AGAGGAAGAG	480
GAAGAAGATT TCGAATCGGG TATTTAAAGG GTAAAAGAAT GATAAAAAAG GTTACGATTG	540
AAAAAATAAA ATCGCCTGAG CGCTTCTTAG AAGTACCACT TCTGACGAAA GAAGAAGTCG	600
GCCAGGCAAT CGATAAGGTT ATTCCGCAGT TAGAACTCAA CCTTGACTAT TTCAAGGAAG	660
ATTTCCCGAC GCCAGCTACC TTTGATAATG TCTATCCAAT CATGGATAAC ACGGAATGGA	720
CCAATGGTTT CTGGACAGGA GAACTGTGGT TGGCTTATGA ATACAGTCAA CAGGATGCAT	780
TTAAAAACAT CGCTCATAAA AATGTTCTTT CTTTCCTGGA TCGTGTCAAT AAGAGAGTAG	840
AATTGGATCA CCATGATCTC GGCTTCTTGT ACACACCGTC TTGTATGGCT GAATATAAGA	900
TAAATGGAGA TGGAGAGGCT AGAGAAGCAA CCTTGAAAGC TGCAGATAAG TTGATTGAAC	960
GCTATCAAGA AAAAGGTGGT TTTATTCAAG CTTGGGGAGA CTTGGGCAAG AAAGAGCATT	1020
ACCGTTTGAT TATCGACTGC TTGCTCAATA TCCAACCTTT ATTCTTTGCT TATCAAGAAA	1080
CAGGCGATCA AAAATACTAC GATATTGCAG AAAGCCATTT CTATGCTTCA GCTAATAATG	1140
TAATCCGTGA TGACGCTTCG TCCTTCCACA CCTTCTATTT TGATCCTGAG ACAGGTCAAC	1200
CCTTTAAAGG TGTAACGAGA CAAGGGTATA GTGATGATTC ATGCTGGGCA CGTGGTCAAT	1260
CATGGGGAGT CTATGGTATT CCTTTGACTT ATCGTCACTT AAAAGACGAG TCCTGCTTTG	1320
ACTTGTTTAA GGGTGTGACC AATTATTTCT TGAATCGTCT GCCAAAAGAT CATGTGTCCT	1380
ATTGGGATTT GATTTTAAAT GATGGTAGTG ATCAATCACG AGATTCTTCA GCAACAGCTA	1440
TCGCCGTCTG TGGGATTCAT GAAATGCTAA AACATCTCCC AGAGGTGGAT GCTGACAAAG	1500
ATATTTATAA ACATGCTATG CATGCCATGC TTCGTTCCCTT GATCGAACAT TATGCAAATG	1560
ATCAATTTAC CCCTGGTGGG ACAAGTCTCC TCCACGGTGT GTACTCATGG CATTCAGGTA	1620
AAGGAGTGGA TGAAGGCAAT ATCTGGGGTG ACTACTATTA CCTAGAAGCC CTTATCCGTT	1680
TCTACAAAGA CTGGAACCTA TATTGGTAGG AGGAGAAATA TGACAATGCC AAATATTATT	1740
ATGACCCGTA TCGATGAACG GTTGATTCAT GGACAAGGAC AACTTTGGGT AAAATACCTA	1800
GGTTGTAATA CGGTCATTGT TGCCAATGAC GAAGTAAGCA CGGACAAGAT GCAACAACT	1860
CTGATGAAAA CAGTTGTGCC AGACTCAGTT GCCATGCGTT TCTTCCCTTT GCAAAAGGTG	1920
ATTGATATCA TTCACAAGGC TAATCCTGCT CAAACGATCT TTATCGTTGT AAAGGATGTG	1980
AAGGACGCTT TAACCTTGGT AGAAGGTGGT GTCATATCA AAGAAATCAA TATTGGGAAC	2040

1191

ATTCACAATG CCCCTGGTAA AGAGCAAGTG ACACGCTCCA TCTTCCTGGG TGAAGAGGAC	2100
AAGGCGGCCC TCAAGGAATT GAGCCAACT CATCAAGTAA CATTTAATAC GAAAACAACT	2160
CCAACAGGAA ATGATGGAGC TGTTCAGTC AACATTATGG ACTATATTTA ACAGAGGAGA	2220
TCGTTATGTC GATTAATGTA TTTCAGCGA TTTTAATTGG ATTATGGACA GCTTTCTGTT	2280
TTAGTGGAAT GCTGTTAGGA ATTTACACCA ATAGATGTAT TGTTCGTCA TTTGGTGTCC	2340
GAATTATTCT AGGTGATCTG TCATGCTCTT GCAATGGGAG CCAATGGTGA ATTGG	2395

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3342 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

CCTTCTTTAG AGGTAAATTT TGCAAAATCG TCGATTGTTA TATAAGGATT ATTATAGAGA	60
CTGTTTCGCA AGAATCTCTG ATATGTTTTT GAATCTTTTG AATACAAAAC TATCTCTCTA	120
ATAGCATTGC CATCTGTTC ATCAATTGGT AAACATACCG TAACTAGAAA AAGAATTATA	180
TTCAAAATAA AAAATTCTGA TCGGTACGGC ACAAATCCCA AAAGTGCTAA TATTGCGACA	240
ATTAGGTTAG CTCCACCTCC CCCAAAGAAG TAGAACACCA AATTCCTATC ACTATTTTTT	300
TCATTAGTAA TGTTTCTATT ACTCATTGA CAATAACCGA ATGCTAATAA CACTGGAAAT	360
TTGAAATATA TTTTTTTTCT GAAATAGAAG AAAAAGGGAG TAGCAAGCAT CTCTAGTTTA	420
TAAGATAAAC ATCTTCCAC TAAAAATGA CCTAGTTCAT GTAATGTAAT TGATATTAAC	480
GAAATTAAAA TCAATCGAAA ATAATAGATT AATGAATCAT TTGGAAAAAT TATCAATAAT	540
AGGAACAATA ACGGAATCAA ACATAAATAT ATGACAGAGT TATTTAATAT TTTCAACATA	600
ATACCATTCC TCTAACTAT TAGCTTCAAA AAGGCGTTTT TTCTCCCAAT ACATCTTCTC	660
AAAATGTTTG GAATCATAAT TTTCTAAAAT TAATTTTATG TCTGGTAAGC TCTTCTTGA	720
TAATCCGTTG TTTTGTACTT AATTTTCCCT TCAAGTACAT CTTCAATTTT ATAAGTTGCC	780
TCCATCAACT GAGCCTCTGC AATATCTTTG AGTGAATTGG TAATTGAAAC TTGGTGTAAT	840
ATCTGTCTCT CCATATATGA AAATATATCT CTAAGATATT CTGACACATT ATCAGAGCCG	900
TTACTCTCAG CAACATCTAA TGTTACAACA AACTTTCCAG CTAATCGAAA AAGATGGCTC	960
CACCCCCAA TCCTTTCAAT AAAGTTTTTT GTGTCCACAG ATACGTTTTG TAAATATACA	1020

1192

GGAGAAGAGA TAATTATAAT ATCAGACTCT AATAACTCTT TTTTATAAC ACCTCCATCA	1080
TCAGCATTAC TTGCGCTATC AATTCCTTTC TTAACAACCT CTTCTGAATC AGAATTAGAT	1140
ATTTCTAGCT CTGAATTGAA AGGTGTCCTG AAAGATATAT CAACATTATT TCTACTAGAA	1200
ATGATACTTG AAAGTCTCTT AGTATACTCT AAAGTCTTAG AGTTATGATT TCGCACTCCT	1260
GCAATATATA ATATTTTATT CATTTTAATT CATCCTCTCA ATTTGAATTT AGTAGATTTT	1320
TCAAGATAGT ATGGTACAAA AACAGACTTT TGTGACTCA CATTATTACA TATGTTTTGT	1380
ATTAAACCAA AATCAATACT ATTTTGGAG TAATTTTGAT TTTAGTTTAA AATCATTTCT	1440
ATAACAGTAG CATATACCTC AAGCCGTTTA GCAATTAGAA TAGAACTTTT CTTTATTATA	1500
TTATTATCTC AACGAAAAGC TACACTATTA AAAATATTTT ATAGAATTAC ATATTAACT	1560
AGTCAATCTT GGTATTTTTA TATTGCTTAA TGAGTGGACA CCTCTATTTT AGAAACAAAA	1620
CTATAAATTA AGCTAGATTT CAAGTAATGA GGGGATAACT ATCTTTTTGT CATCTGATT	1680
CAGTGGGATA TACCTTAAAA AAGTATAAGC AATACCAGTC ACACCTGTAT ACAAAGAAAA	1740
ATCTGGGAAA TTGCTTGTTT GGACGATACG ATACTCTCCT TCTTTTGATT TATTCATTAC	1800
AACACTACAC AATAAAGACT CCAATTCCAT ACTAGTATCC ATTTCTTTCA TGTAGTCGAT	1860
GTAAAAATTT ATTATGGCCA TACTTCCATG GCAAAATGTA TCATTATCTA AACTAGCTAC	1920
AATTCCTCTT GGAACACTTT GGGGATGATT AACTAATGTC CCAAATCTC CACTACACCA	1980
CTTCAAAGAA TGAATTTTGA TTTTCTCCCT AGGAACTAGT TGTAAAAATTA ATTCTTTATA	2040
TTTTTTAAGT CTTGTCACTT TATAAATATT TTTTAATGTA AAAATTACAC CTGATAGTCC	2100
ATGGCCAAAA CTATATCCAA AATTACTATT ATCTCTCTCG CTTACATCAT TATATAGCGT	2160
ATCACCTAAA CTTAATACTA GCCTTAGAAC ACGTTCCTTC TCTATTCCTC TCCTATAATA	2220
TCTTACCAGT GTATTAAATTA AAGGTAGAAG ACCATTAATA TAGTCAGACT TGTTTGAAAC	2280
ACTTGCAAAA TCAGTCTTTT CAAGCTCAGT TAAAACACTC TTTATATAAT TTAAGCATGC	2340
GAGAGTATTT GTATCGTAAT CCTCTATAAT GGATAGAACA ATGAAATATC CTATATCCCC	2400
AGTTAAACCA AATGTGGTCT TAGATAAAGA AACAGATGGC GGAATTGCAG ATAACATTTT	2460
ATTGTACAGT TGAGTATATG ATGATTTATC TTTCAATAAT TTTACATAGT ACATAAACAG	2520
TAATATTCCA GCTCTACCCC TATACATATC ATTTCCCGTT TGTCAAGAC ACCATTTAGA	2580
ACCTTTAAAA TTAACAGGTA TACTCCAAAT TGGATATTCG TCATAAATAT TATTAATAAC	2640
CAAAGAGTCT GCAATATTTT CTACTTCATT ATGCAGAATA GTAACATAAC TTTCATTTGG	2700
GAGTTTTTTT CTATTAGATA AGTTTAATTT ATATCCTTTT TTTGCTGAT CAAAGCTTGG	2760
AAAATAAATT TCAATGATAT CAAGTTGCTT TTCTAAATTT TCCAAATTAT TATTAGGTAA	2820



1193

ATATTTTCATA AAATAGTCAT ATCCAGAAAA TTGATGTAGG GAAATAAAAT GATTTCCAAA	2880
ATCATCGTAG ATTTTCATTGA TATTTGTATC TGTATAAAAA ATCGGAATAT CTAATAACCT	2940
CATTGTGTTCA CATTGCTTG CTACAATACC TTGATTAGAA AACTTATTGC TCCAGAGATT	3000
TTCCAATGCT TTTTCTCTAT CTAACATTTT TCATAAAAA TCAGGATGAT ATAAAAAGA	3060
TAGTACTGAA GCATAGCTAT TTGTGTCTCT AAAAAGTACC CTTGTCTTTA AACCATACAA	3120
GTTTGCTTTT AATAGCATTT TAAATTCTTC TGTTTTATTT AACTCTTCAA ATATCAGATA	3180
AAAATCCCTA AAACCTTTTT TGAAATCTTT TATATACTTA TCAAATTCTA TATCACCATC	3240
CCGAACAGGC AGGTTTTTCC CACCTTCAAA ATCAATTTTC CCAATATCAA ACTTTACCTT	3300
ATCAGTATTT AAATTAATTA AAACCTGACC AGGGATCCTC TA	3342

(2) INFORMATION FOR SEQ ID NO: 207:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3454 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

GAGAAAAGAA TGTTAAAGAA AAATGATATT GTAGAAGTTG AAATTGTTGA TTTGACCCAT	60
GAAGGGGCAG GAGTTGCCAA GGTAGATGGT TTGGTCTTTT TTGTAGAGAA TGCTTTACCG	120
AGTGAAAAAA TTCTCATGCG TGTCCTCAAG GTCAATAAAA AGATTGGCTT TGGAAAAGTT	180
GAAAAATACC TTGTCCAGTC ACCACACCGT AATCAAGATC TAGATTGGCC TTACCTGCGT	240
TCAGGAATCG CGGATTTAGG ACACCTTTCT TATCCAGAAC AGCTCAAGTT TAAACCAAG	300
CAAGTCAAGG ACAGTCTCTA CAAGATTGCT GGAATTGCAG ATGTAGAAGT TGCTGAAACG	360
CTTGGTATGG AACATCCAGT CAAGTATCGC AATAAGGCGC AGGTGCCCGT TCGTCGAGTG	420
AATGGTGTCT TGGAAACAGG ATTTTCCCGT AAGAATTCCG ATAACCTCAT GCCCCTTGAA	480
GATTTCTTTA TCCAGGATCC TGTCATTGAC CAAGTCGTAG TAGCTCTTCG AGACCTGCTC	540
CGTCGTTTTG ATTTAAAACC TTATGACGAA AAGGAACAGT CTGGATTGAT TCGGAATCTT	600
GTGGTGCGTC GTGGTCACTA TTCAGGACAA ATCATGGTCG TTTTGGTGAC AACTCGTCCA	660
AAAGTTTTTC GTGTTGACCA ATTGATTGAA CAAGTTATCA AGCAGTTCCC AGAGATTGTG	720
TCTGTCATGC AAAATATCAA CGACCAGAAT ACCAATGCCA TTTTGGTAA GGAGTGGCGC	780
ACTCTTTATG GTCAAGACTA TATTACGGAC CAGATGTTGG GAAATGACTT CCAAATCGCT	840

1194

GGCCCAGCCT TTTACCAAGT CAATACTGAA ATGGCCGAGA AACTCTATCA AACAGCCATT	900
GACTTTGCAG AGTTAAAAAA AGATGATGTG ATTATTGATG CCTATTCTGG TATTGGAACC	960
ATTGGTTTAT CAGTCGCCAA GCATGTCAAA GAAGTCTACG GTGTTGAACT GATTCCAGAA	1020
GCAGTAGAGA ATAGCCAGAA GAATGCTTCT TTGAACAAGA TTAATAATGC CCACTATGTC	1080
TGTGACACGG CTGAAAATGC CATGAAGAAA TGGCTCAAGG AAGGTATTCA ACCAACCGTT	1140
ATCTTGGTTG ATCCTCCACG CAAGGGCTTG ACAGAAAGCT TTATCAAAGC AAGCGCCCAA	1200
ACAGGAGCCG ATCGCATCGC CTATATCTCC TGCAATGTCG CAACCATGGC GCGTGATATT	1260
AAACTATACC AAGAGTTGGG ATATGAATYG AAGAAAGTCC AGCCGGTGGA TCTATTTCTT	1320
CAAACGCATC ACGTCGAGAC GGTAGCACTT TTGTCCAAAC TCGATGTCGA TAAGCACATA	1380
ACTGTTGAAA TTGAGCTGGA TGAGATGGAT TTGACAAGTG CGGAGAGCAA AGCAACATAT	1440
GCTCAAATCA AAGAATATGT TTGGAATAAA TTTGAATTAA AAGTTTCGAC ATTATATATT	1500
GCACAGATAA AAAAGAAATG TGGAAATAGAA TTACGAGAAC ATTACAACAA GTCTAAAAAG	1560
GATAAACAAA TTATTCCACA GTGTACACCT GAAAAAGAAG AAGCCATCAT GGATGCTTTG	1620
AGACACTTCA AAATGATTTA ATAGAAAAGA ATGACAGTAT ATGACTTTCT GCATTTATTA	1680
CATTCCTACT TGGTATAGGA ACAGCTATTA TTCCTTTCTT GCAAGGTATC AATTAGAAAA	1740
TAGGCTCAAT ATAAAGATTG ATAGGATCAT TTTTATATTT AAAGGAGCGT TGAATGATT	1800
GATAAAGGCA ACAAAAAATT TTAGGATAAA TTTGCTAAGT TGTATGCCTC TTTTATGAAA	1860
AAAGATAAAG AGGTTTATGA TAAAGTTTGT GAATATCTTA GTCCTCATTT GAATAAAGAT	1920
ATGGAGGTGC TTGAACTTGC TTGTTGGTTT CGTGTCAATA CAGTTATAGA GGCAAATAGT	1980
TATGTAAATA TAAGGAGTTC AAGACTTCTA CCAAAGTTTA AAACCTCAAAA AATAAATAGT	2040
TGGTGTGCTG CTTACAATAT CCATTTTAAT AATGGATATT GTAAGCAGCA CCCCcAtGAA	2100
TTTAAAGATT CTTTAAAGAG TCTTATTTTG TGATGAAAA TTAATATGTA AATCTCAGAC	2160
GATAGAAATT AAAAATCTA TCGTCTTTTT TATACTCAA ATTAGGAGGT AAAAATGGTA	2220
AGGATAAGAG GTCCCACCTA AAACAATTTA TGGCAAAATA AGGACGGAAT AACACAACAA	2280
ATTCTCTAAA ACAAATCACT AAATCAATGT AAGATTGAAT GAAATCAATA TTTATGCTAT	2340
AATTAAATAA ATTTAATGAA GAAAAAAGA GGGATATTAT GGCACCTAAC TATAAACCAT	2400
TATGGATACA GTTAGCAAAA AAAGGACTAA AGAAAACAGA TGTAATAGCT ATGGCAGGAC	2460
TTACAACAAA TGTATGGCA CAAATGGGAA AGGATAAACC AATTACATTT AAGAATTAG	2520
AAAGAATATG TAAGGCTTTA TCTTGCACTC CTAATGATAT TATTAGTTTT GAAGATAATT	2580
TTAGTGACGA GGAATAGAAA ATGACTTTAA GGACAGAAGA TCAAGTTAGG GATTATGCAA	2640

1195

GAGAAGTATA GGCTTTAATG AAGTTGAAGA AAACATCAAT CAAGGTACTG GTCAAATAAC	2700
TACTTTTAAAT CAATTAGGCT TCAAGGGATA TTCAAATAAG CCAGATGGTT GGTATTTACC	2760
TAAAAATATG AATGATGTAG CAATAATCCT TGAAACAAAA TCAGAAGAAA GAGATATTAG	2820
CAAACAAATT TTTATTGATG AGTTAATGAA AAATATAGAC ATAATTTAAC TAAAAATAAA	2880
AACTAGATCC TTTTTTGAAA AAATTATATT ATTAAATTG TAACTGTATC TATTGACAAT	2940
<hr/>	
GATAATTATT ATCGATACAA TAGACTTGAA ATATGTTTAA GGAGTTTTTA TGAAAAAATAA	3000
TTTTTCTAA TmGCTATTTT AGCTATGTGT ATAGTTTTTA GCGCTTGTTT TTCTAATTCT	3060
GTTAAAAATG AAGAAAATAC TTCTAAAGAG CATGCGCCTG ATAAAATAGT TTTAGATCAT	3120
GCTTTCGGTC AAATATATT AGATAAAAAA CCTGAAAGAG TTGCAACTAT TGCTTGGGGA	3180
AATCATGATG TAGCATTAGC TTTAGGAATA GTTCTGTTG GATTTTCAAA AGCAAATTAC	3240
GGTGTAAGTG CTGATAAAGG AGTTTTACCA TGGACAGAAG AAAAAATCAA AGAACTAAAT	3300
GGTAAAGCTA ACCTATTTGA CGATTGGAT GGACTTAACT TTGAAGCAAT ATCAAATTCT	3360
AAACCAGATG TTATCTTAGC AGGTTATTCT GGTATAACTA AAGAAGATTA TGACACTCTA	3420
TCAAAAATTG CTCCTGTAGC AGCATACAAA TCTG	3454

(2) INFORMATION FOR SEQ ID NO: 208:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3752 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

CGGGAGTATA CTTAATATAA TTATAGTCTA AAAATGACTA TCAGAAAAGA GGTAAATTTA	60
GATGAATAAG AAAAAAATGA TTTTAACAAG TCTAGCCAGC GTCGCTATCT TAGGGGCTGG	120
TTTGTTACG TCTCAGCCTA CTTTGTGAAG AGCAGAAGAA TCTCCAAG TTGTCGAAAA	180
ATCTTCATTA GAGAAGAAAT ATGAGGAAGC AAAAGCAAAA GCTGATACTG CCAAGAAAGA	240
TTACGAAACG GCTAAAAGA AAGCAGAAGA CGCTCAGAAA AAGTATGAAG ATGATCAGAA	300
GAGAACTGAG GAGAAAGCTC GAAAAGAAGC AGAAGCATCT CAAAATTGA ATGATGTGGC	360
GCTTGTGTT CAAAATGCAT ATAAAGAGTA CCGAGAAGTT CAAAATCAAC GTAGTAAATA	420
TAAATCTGAC GCTGAATATC AGAAAAAATT AACAGAGGTC GACTCTAAAA TAGAGAAGGC	480
TAGGAAAGAG CAACAGGACT TGCAAAATAA ATTTAATGAA GTAAGAGCAG TTGTAGTTCC	540

1196

TGAACCAAAT GCGTTGGCTG AGACTAAGAA AAAAGCAGAA GAAGCTAAAG CAGAAGAAAA	600
AGTAGCTAAG AGAAAATATG ATTATGCAAC TCTAAAGGTA GCACTAGCGA AGAAAGAAGT	660
AGAGGCTAAG GAACTTGAAA TTGAAAACT TCAATATGAA ATTTCTACTT TGGAACAAGA	720
AGTTGCTACT GCTCAACATC AAGTAGATAA TTTGAAAAA CTTCTTGCTG GTGCGGATCC	780
TGATGATGGC ACAGAAGTTA TAGAAGCTAA ATTAAAAAA GGAAGAGCTG AGCTAAACGC	840
TAAACAAGCT GAGTTAGCAA AAAAACAAAC AGAACTTGAA AAACCTTCTG ACAGCCTTGA	900
TCCTGAAGGT AAGACTCAGG ATGAATTAGA TAAAGAAGCA GAAGAAGCTG AGTTGGATAA	960
AAAAGCTGAT GAACTTCAAA ATAAAGTTGC TGATTTAGAA AAAGAAATTA GTAACCTTGA	1020
AATATTACTT GGAGGGGCTG ATCCTGAAGA TGATACTGCT GCTCTTCAAA ATAAATTAGC	1080
TGCTAAAAA GCTGAGTTAG CAAAAAACA AACAGAACTT GAAAACTTC TTGACAGCCT	1140
TGATCCTGAA GGTAAGACTC AGGATGAATT AGATAAAGAA GCAGAAGAAG CTGAGTTGGA	1200
TAAAAAAGCT GATGAACTTC AAAATAAAGT TGCTGATTTA GAAAAAGAAA TTAGTAACCT	1260
TGAAATATTA CTTGGAGGGG CTGATTCTGA AGATGATACT GCTGCTCTTC AAAATAAAT	1320
AGCTACTAAA AAAGCTGAAT TGGAAAAAC TCAAAAAGAA TTAGATGCAG CTCTTAATGA	1380
GTTAGGCCCT GATGGAGATG AAGAAGAAAC TCCAGCGCCG GCTCCTCAAC CAGAGCAACC	1440
AGCTCCTGCA CCAAAACCAG AGCAACCAGC TCCAGCTCCA AAACCAGAGC AACCAGCTCC	1500
TGCACCAAAA CCAGAGCAAC CAGCTCCAGC TCCAAAACCA GAGCAACCAG CTCCAGCTCC	1560
AAAACCAGAG CAACCAGCTA AGCCGGAGAA ACCAGCTGAA GAGCCTACTC AACCAGAAAA	1620
ACCAGCCACT CCAAAAACAG GCTGGAAACA AGAAAACGGT ATGTGGTATT TCTACAATAC	1680
TGATGGTTCA ATGGCAATAG GTTGGCTCCA AAACAACGGT TCATGGTACT ACCTAAACGC	1740
TAACGGCGCT ATGGCAACAG GTTGGGTGAA AGATGGAGAT ACCTGGTACT ATCTTGAAGC	1800
ATCAGGTGCT ATGAAAGCAA GCCAATGGTT CAAAGTATCA GATAAATGGT ACTATGTCAA	1860
CAGCAATGGC GCTATGGCGA CAGGCTGGCT CCAATACAAT GGCTCATGGT ACTACCTCAA	1920
CGCTAATGGT GATATGGCGA CAGGATGGCT CCAATACAAC GGTTCATGGT ATTACCTCAA	1980
CGCTAATGGT GATATGGCGA CAGGATGGCT TAAAGTCAAC GGTTCATGGT ACTACCTAAA	2040
CGCTAACGGT GCTATGGCTA CAGGTTGGGC TAAAGTCAAC GGTTCATGGT ACTACCTAAA	2100
CGCTAACGGT TCAATGGCAA CAGGTTGGGT GAAAGATGGA GATACCTGGT ACTATCTTGA	2160
AGCATCAGGT GCTATGAAAG CAAGCCAATG GTTCAAAGTA TCAGATAAAT GGTACTATGT	2220
CAATGGCTTA GGTGCCCTTG CAGTCAACAC AACTGTAGAT GGCTATAAAG TCAATGCCAA	2280
TGCTGAATGG GTTTAAGCCG ATTAAATTAA ATCATGTATA GAACATTGGA CATTTTAATT	2340

1197

TTGAAACAAA GATAAGGTTT GATTGAATAG ATTTATGTTT GTATTCTTTA GGTACCTATC	2400
TTATGATTTT AGGAAATGTC ATTAAAAAAA CGACTCATTT TCTCTAACCT GAAAAATAGA	2460
TTAGAGAAAA TGGGTTGTTT TATCTATTAT AGTTATTTGA ATGAAGmTAA GAAGAAGGTA	2520
TACTCACATC ATTCACATAA TCTGTATATT GACTATAAGT TTAAAAAAC AATTTTAAAG	2580
CTCTTCCTTG TCTTCTCTAA CCAAGCGTGT TATAATGAAT ACTGCTCAAG CGACCTTCAA	2640
TCGTGAAGCA CACACGACCT TCAATCGTGA ATAAACGAAT AGATGGGAGA CTTACCATGA	2700
GTGATAACTC TAAACACGT GTTGTCTGGG GGATGACTGG TGGTGTGAT TCGTCGGTGA	2760
CGGCTCTTTT GCTCAAGGAG CAGGGCTACG ATGTGATCGG TATCTTCATG AAGAACTGGG	2820
ATGACACAGA TGAAAACGGC GTCTGTACGG CGACCGAAGA TTACAAGGAT GTGGTTGCGG	2880
TGGCAGACCA GATTGGCATT CCCTACTACT CTGTCAATTT TGAAAAAGAG TACTGGGACC	2940
GCGTTTTTGA GTATTTCTTA GCGGAATACC GTGCAGGGCG CACGCCAAAT CCGGACGTTA	3000
TGTGCAACAA GGAAATCAAG TTCAAGGCCT TTTTGGACTA TGCCATAACC TTGGGGGCAG	3060
ACTATGTAGC GACTGGGCAT TATGCTCGAG TGGCGCGTGA TGAGGATGGT ACCGTTTACA	3120
TGCTTCGTGG CGTGGACAAT GGCAAGGATC AGACCTATTT CCTCAGCCAA CTTTCGCAAG	3180
AACAACCTCA AAAAACCATG TTCCCACTAG GACATTTGGA AAAGCCTGAA GTACGCAGAC	3240
TAGCAGAAGA AGCAGGCCTT TCGACTGCTA AGAAGAAAGA CTCGACAGGG ATTTGCTTTA	3300
TCCGAGAAAA GAACTTTAAA AACTTTCTCA GCAACTACCT GCCAGCTCAG CCTGGTCGCA	3360
TGATGACTGT GGATGGTCGC GATATGGGCG AGCATGCAGG TCTTATGTAC TATACAATCG	3420
GTGACGCTGG CGGACTCGGT ATCGGTGGGC AACACGGCGG TGACAATGCC CCTTGGTTTCG	3480
TTGTCCGAAA AGATCTAAGC AAGAATATTC TCTATGTAGG ACAAGGATTC TACCATGATT	3540
CGCTCATGTC AACTAGCCTA GAAGCCAGTC AAGTCCACTT TACTCGTGAA ATGCCAGAAG	3600
AGTTTACGCT AGAATGTACG GCTAAATCC GTTACCGTCA GCCTGACTCT AAGGTGACCG	3660
TTCATGTCAA AGGAGAAAAG ACAGAGGTCA TCTTTGCCGA ACCACAACGC GCGATTACAC	3720
CAGGACAGGC AGTTGTCTTT TACGATGGCG GG	3752

(2) INFORMATION FOR SEQ ID NO: 209:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 3580 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1198

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

TATTTATATT TTTTATCTC TGGCATACTT TGATACCTTT TTAGACTTAA AGTCTTTAAT	60
AGTGCCTTTC CACCTCTTTT TATCTATAAA GATTCTCCTA CATCATAATT CATTTTMTTA	120
TTTAAACCTT TCTGCTTAG TTTGTCTTTA TCTTCTTCAT ACCATTTTAA GATTGTCACA	180
TAGTGGTTTT GATAGGTCTT ACCACTGCTT TCCATGTATC TGGATAGTTT ATTTATCATT	240
ATATCTGTGT GTGAGTTTAA TTTTCTTTT AGATTTTAT ATTCTTCTT GCTTAACCTT	300
ACATTTTGA ATTCTCCATA AAAAATGGGG GTGGACTTTT TATCTATCTC TCCCTCTCTC	360
TCTTTATCTA TCTCTATATC TTTCCATGTA ATTCCAATCT GGAGTACCTC TACTGTCTAT	420
CGGTAATTTA ATTTTGATAT CTGGCAATAC TGTGCTAGAT ATTTGATCTT TATATTCACT	480
ATTTTAAAA GCTTGCCTAA TAATTGAAGT TAAATAGAAT GCTACTTCTT TATTCAATTC	540
TTTATTTTTT AATTTTAAAC AATGAATTTT CATATCTAGG CTGCTTTAT ATTTATGATA	600
AAAGACTGCT CCTAAAAATG AAACAGATAT AAAATTTTCA AAAACTCTAT AATTTTATC	660
ATCTATATCT TCGTAGTAAC CTAAGATACC ATTGTCAATA TTTGTAGCAC TAATTCTAGG	720
AGTTTTTCCA TCGAGTAAAT ATCTTTTGG AATAGATGAG CCTGTTGGTA CTTAACTCGA	780
TTTCCCTTTT TTTTCGGTAA TAAATATTTT TTTTATTTT GTGCTCTGAT ATTTTCTTA	840
CCTGCTCTTT GTAGGATGAG TATTTTCTAG ATTTTCyTGA ATAACCTTTT ACTGAAGTT	900
TTAGCTTTTG AACTAGTCGT TGTACTTTCT TTTTGTATAT TATCAGTCCT GATCTTTTAA	960
ATATTGCTGT TATTCTCTAT ATCCTATTTT TCATTCAATGA TATTCTTTTA CTAATTTTAT	1020
CTTAAATTCT GTGCTGTATT TGCCATTAAT AAACTGACCT CCTTAGTTA GTTTTTTGGC	1080
CTAACTTTTG AGGGTCAGTT CAAAATTTGC GACTTTTAAA TGAATTCCAA TATTCAATTA	1140
TTAAGAGTTA ACATGGTGCT TGCCAATAGG AATCATTAGA GGCGAATTGG AAATAGGGTC	1200
ACGTATAATT TTTGCTTCAA GATTAAAGAT ATCTTTAACT AGTTTATCAT TTAGTATATC	1260
TTCAGGCTTT CCCTCTGCAA CAAGTTTACC TTCTTTAATT GCAAATAGGT AATCAGCGTA	1320
TCTTGCTGTT AGATTTATAT CGTGCAAAAT CATGCAAAATG GTTGCTTTAT ATTTTGGTT	1380
TAGATCAGTC AAGAGGTCTA ATAGTTCTAT TTGATATGAG ATATCCAAGT AAGTAGTTGG	1440
CTCATCTAAA AGTAGGATAC TTGTATCTTG GGCTAGGGCT AGAGCTATCC ATACTCTTTG	1500
CCTTTGACCC CCAGAAAGTT CTTCAACTAG GTTATTTGCT AGATCTTCAA CATTGGCCTT	1560
AACCATTGAT CTGTTTATTA TTTCAAGGTC ATCTTTTCCA AGACTCTTAA AAGGCTTTCT	1620
GTAGGGGAAA CGACCACGGC TTACAAGATC AGCTACTGTT ATTGATTCAG GGATTATTGG	1680
AGATTGAGGT AATATAGCTA TGTGTTTTGC TAAATCTTTT TCTTTATAAG AATTAATTGA	1740

1199

TTTATTATCA AGCAATACTT CTCCCTCTAA TGGCTTTATA AGTCGAGACA AGGTTTTAAT	1800
GAGTGTTGAT TTCCCACAAC CATTTGACCC AATAATAACT GATATTTTTT CTTCAGGTAT	1860
TTTTATATTT ATATTTTCCA AGATTATTTT TTCATCATAA CCGCAGGTAA GATTATTTGA	1920
CCACAGACCT TTCATTATAT ATTCTCCTG TTCATTTTTA TTAGTAAGTA TATTAAGTAT	1980
GGTGAACCTA ACAAGCCAGT TACAACACCT ACTGGATATC TAGCTGGTAA AATATTTTGA	2040
GAGAATATGT CTGATAACAA AACTAGTAAA ATTCCAACCA ATCCAGCTAA TATTGGGCTT	2100
CTTTCTTGC CAATATTTAA GGCTATGGGA CCAGCTAAAA AAGATATACA AGCTATTGGT	2160
CCTGTAATTG AAGTAGAAAA AGCAGTTAAA GATACAGCGC AAAAAATTAA AACAAGCCTT	2220
GAAAGCTCGG GATTTGCTCC AAGTCCGATT GCTATTTCTT CACCAAGTTC AATAATTTCT	2280
AGTCTTTTAT TAAAAAATAA AACTAATATA GTAGCAATAA TACTTACTAT TAGAACAAGA	2340
GGTATGTCAT CTAACTTTGT AAAAGATAAA GAGCCACTGA GCCATCTCAT AACTTCTTGT	2400
AATTCATATC TTGCTACTTT CAACAATAAA AATGAGGTGC CTGCTCTTGT GACAGCTTGA	2460
AAACCAATAC CTAATATTAT CAGTCTTGCT GCTGAAAAAC CATCTTTTTT AGCTAGTAAA	2520
AATAATATTA AAGATGATGT TAGTCCACAA GTTATTGAAA TAATTCAGT AGTTAAACTA	2580
TTTGTTTTTA ATACCAATAT GCAAAAGACC GCTGCAATAG ATGAAGAACT TGTGACACCG	2640
ATTATATCAG GACTTGCAAG AGGATTTCCT AACATAGTTT GAAAGATAAA TCCTGCCAAT	2700
CCAAAAGACC AGCCAGCTAT AATTCCTGCT AATAATTTTG GTAATCTAAT TTCCATAATC	2760
GAAAACTAG CTCCAGGAAC AGTTTCACTA TTTAAGACTT TAATCAAAGT TGAAAAAGAA	2820
TAACTTTCAT CTCCGATAAG TAAAATGAAA AATGATAGAC TGATTATTAT TAATAAAAAAT	2880
AGTGAGGAAA ATAGTGTTAT TCTATTTTTT CTTTTTTGAA TACCTATAAT TAAATTTTGC	2940
ATTAGTTATT AACCCCTCTA TTTTTCATAG TTACATAAAT AAGTACTGGA CCCCCGATTA	3000
TTGCAGTAAT TATCCCTACT TCAATTCAC CTGGTTTACC TAACATACGG CCGATTATAT	3060
CACATATAAG CAAGAGCTCT GCACCTATAA AAGATGAAGA AATGGTCATT GTGCGTATAT	3120
CTTTGCTTAT AAATAAGCCA CAAAAGTGAG GAACTATAAG ACCTACGAAG CCAATAGGTC	3180
CACCAATTGC AGTAATACTT GAACATAAAA GCACACTTGC AATTATTGCA AGTGATCTTA	3240
TCCTATTAAC ATTAACCTCA AGACCAACAG CCATTTTCATC ACCCATAGCT AAAGCGTTTA	3300
AATCTGATGA AATAAATATA GCTATCAAGT GACCTAAAAT TATAAAAGGT AGTAGTGTAG	3360
ATATAGAAGA TAATGTAGCT GCTCCAAGGC TACCTATTTG CCAAAATCTA AATTTGTCTA	3420
AGACGTTATT ATTCGGTAAA ATTAAAAAAC TTACAAAACCT GCTTAAAGCC ATACTAACAC	3480

1200

AAGTTCCTGA TAAGGCAAGT TTTATAGGGG TAAGGCCTGC TTTCCGTTA CAGCAATCGC	3540
GTATACAAAA ATTGCACTTA CTAAGCCACC AATGATTGCG	3580

## (2) INFORMATION FOR SEQ ID NO: 210:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 11378 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

CCAAATTGCT CCACAATTAT TATGGAGTCG TCGTTTGGCA GATGGGCGTG ATATGTGTGC	60
TCAAGAATGG TTGACAGGCA AGATATTGAC CCCCTATGAT ATGAATCGTA AGCAAATCGT	120
CAATATTTTA ACCCGTCTTC ATCGCTCAGC TCCGTTGATG ACACAATTGA GTCGTTTGGG	180
CTATGCCATG GAAACACCTG TAGATTTACT ACAGTCTTGG CAGGAAACGG CTCCAGATGC	240
TTTGGGTAAA AATCATTTTA TCAGTGAAGT GATGGCTGAT TTACGTCAGA CTATTCCAGG	300
ATTTAGAGAG GACCATGCGA CCATTGTCCA TGGAGATGTA CGACATAGTA ATTGGATTGA	360
GACAGATAGT GGCTTGATTT ATTTAGTAGA TTGGGATTCG GTTCGCTTGA CCGATCGCAT	420
GTTTGATGTG GCCCATATGC TCTGCCATTA TATTTACAGAA CATCAGTGGG AGGAATGGTT	480
GACCTACTAC GGTTACAAGT ACAATCAAAC GGTATTAAGT AAATTGTATT GGTATGGTCA	540
ATTGCTCTAT TTGAGTCAGA TTTCCAAGTA TTATATGAAC CAAGATTTAG AAAATGTCAA	600
TCGGGAGATT CATGGTTTGC GTCATTTCCG AGACAAGTAT GGAAAGAGAA GATGAGAGTT	660
AGAAATCGTA AAGGGGCAAC AGAATTACTA GAGGCAAATC CCCAGTATGT GGTCTCTAAT	720
CCCTTGGAAG CCAAGGCAAA ATGGCGGGAC TTGTTTGGCA ATGATAATCC CATTATGTG	780
GAAGTTGGAA GTGGAAAGGG TGCCTTTGTT TCAGGTATGG CCAAGCAAAA CCCTGACATC	840
AACTATATCG GGATTGATAT TCAAAAGTCT GTTTTGAGCT ACGCTTTGGA CAAGGTGCTT	900
GAAGTTGGAG TGCCTAACAT CAAGCTCTTG TGGTAGATG GTTCTGACTT AACTGACTAC	960
TTTGAAGACG GTGAGATTGA TCGCTTGAT CTGAACTTTT CAGATCCATG GCCGAAAAAA	1020
CGCCATGAAA AGCGTCGTTT GACCTACAAG ACCTTCTTGG ATACCTTCAA ACGTATCTTG	1080
CCTGAAAATG GAGAAATTCA TTTCAAGACG GATAACCGTG GCTTGTTTGA GTACAGTTTA	1140
GTGAGCTTTT CTCAATATGG CATGAAATC AATGGTGTCT GGTAGATTT GCATGCCAGT	1200
GATTTTGAAG GCAATGTCAT GACAGAATAC GAGCAAAAAT TCTCAAACAA GGGGCAAGTT	1260
ATCTACCGAG TTGAGGCAGA ATTTTAAGAG ATAACTAAA ATTAGGCTGT ACAAGTGCTT	1320



1201

TTGCTTTACA TAAGTTGGCA AACGTGCTAT ACTGATAGTA AGAATATGAA AAGTGAGGCG	1380
GGGAAATATC TTCGCCTCTT GCTTATGAGG AGGTGGACGC AATCGCAACA ATCGTAGAAT	1440
TAGTCAGAGA AGTTGTAGAA CCTGTCATAG AAGCTCCTTT TGAAC TCGTG GATATCGAGT	1500
ATGGAAAGAT TGGCAGTGAC ATGATTCTCA GTATTTTGT AGATAAACCC GAAGAATTAC	1560
CTTGAACGAC ACGGCAGACT TGACAGAAAT TATCAGTCCT GTCCTAGACA CCATCAAGCC	1620
AGATCCCTTC-CCAGAACAAT ATTTCTTAGA AATTACCAGT CCAGGTTTGG AACGTCCTTT	1680
GAAAACCAAG GATGCCGTCG CTGGAGCGGT TGGAAAATAC ATCCATGTCG GGCTCTACCA	1740
AGCCATCGAT AAGCAAAAGG TCTTTGAAGG AACCTTGTG GCCTTCGAAG AGGACGAGTT	1800
GA CTATGGAA TATATGGACA AGACGCGTAA GAAAACCGTC CAAATTCAT ACAGTTTAGT	1860
ATCAAAAGCA CGTTTAGCAG TTAATTATA GAAAAGAAA GGATAGCTTT TGAGGATTCA	1920
AAAGTGAAGA AAACATGACT AAAGAAATGC TAGAGGCCTT CCGCATTTTG GAAGAAGACA	1980
AGGGAATCAA AAAAGAAGAT ATCATCGACG CAGTAGTAGA GTCGCTTCGT TCCGCTTATC	2040
GCAGACGCTA TGGTCAGTCA GACAGCGTAG CTATTGACTT CAACGAAAAA ACAGGTGACT	2100
TTACAGTTTA TACTGTCCGT GAAGTTGTTG ATGAAGTATT TGATAGCCGT TTGGAATCA	2160
GCTTGAAAGA TGCTCTTGCC ATTAATTCAG CTTATGAACT TGGAGACAAA ATCAAGTTTG	2220
AAGAAGCACC AGCTGAGTTT GGTGCTGTAG CAGCCCAATC TGCCAAACAA ACCATCATGG	2280
AAAAATGCG CAAGCAAACA CGTGCCATCA CTTACAATAC TTACAAAGAA CATGAGCAAG	2340
AAATCATGTC TGGTACAGTA GAACGCTTTG ACAACCGCTT TATCTATGTC AACCTTGGA	2400
GCATCGAAGC CCAATTGTCA AAACAAGACC AAATTCCTGG AGAAGTTTTT GCTTCTCATG	2460
ATCGTATCGA AGTTTATGTT TACAAGGTTG AAGACAACCC TCGTGGTGTG AACGTCCTTG	2520
TTAGCCGTAG TCATCCAGAA ATGATCAAAC GTTTAATGGA GCAAGAAATT CCAGAAGTTT	2580
ATGATGGAAC TGTTGAAATC ATGAGCGTGG CTCGTGAAGC AGGTGACCGT ACGAAGGTTG	2640
CTGTTCTAG CCACAATCCA AACGTGGATG CTATCGGTAC AATCGTTGGA CGTGGTGGTG	2700
CTAATATCAA GAAGTTACT AGCAAATTC ACCCAGCTCG TTACGATGCT AAAAATGACC	2760
GCATGGTACC AATCGAAGAA AATATCGATG TTATCGAGTG GGTAGCAGAT CCAGCTGAAT	2820
TTATCTACAA TGCCATCGCT CCTGCTGAGG TTGACCAAGT TATCTTTGAT GAAAACGACA	2880
GCAAACGTGC CTTGGTGGTT GTTCCAGATA ACAAGCTTTC TCTTGCCATT GGTGCTCGTG	2940
GACAAAACGT GCGCTTGGCG GCTCACTTGA CTGTTTACCG TATCGATATC AAGTCTGCTA	3000
GCGAATTTGA AGCCATGGAA GACGCTGCTT CAGTAGAGTT GGAAGTAGAA AACGATACTG	3060

1202

TAGAAGAATA	AAAGCTGCTA	GAGGAGGGAA	AGATGAAAAAC	AAGAAAAATC	CCTTTGCGCA	3120
AGTCTGTTGT	GTCTAACGAA	GTGATTGATA	AGCGTGATTT	GCTCCGCATT	GTCAAGAACA	3180
AGGAAGGACA	AGTCTTTATT	GATcCTACGG	GCAAGGCCAA	TGGCCGCGGC	GCTTATATCA	3240
AACTAGACAA	TGCAGAAGCC	CTAGAGGCCA	AAAAGAAGAA	GGTCTTTAAC	CGCAGCTTTA	3300
GCATGGAAGT	GGAAGAAAGC	TTTTATGACG	AGTTGATCGC	TTATGTGGAT	CACAAAGTGA	3360
AAAGAAGAGA	GTTGGGACTT	GAATAAGCAA	AAGATAAGTA	ATCTCTTGGG	GCTTGCTCAG	3420
CGAGCAGGGC	GCATCATATC	GGGTGAAGAA	TTGGTGGTCA	AGGCCATTCA	AGACGGCAAG	3480
GCCAAGTTGG	TCTTTCTAGC	TCATGATGCT	GGACCCCAATC	TGACCAAGAA	GATTCAAGAT	3540
AAAAGTCATT	ATTATCAAGT	AGAAATTGTA	ACCGTGTTTT	CAACACTGGA	ATTAAGCATA	3600
GCAGTCGGGA	AATCGAGAAA	GGTTTTGGCT	GTAACAGATG	CTGGATTTAC	AAAGAAAATG	3660
AGGTCTCTTA	TGGAATAGAA	GAGGAGGACA	TGATTTGTCT	AAGAAAAGAT	TGTACGAAAT	3720
CGCAAAAGAA	CTTGGAAAAG	AAAGTAAAGA	AGTTGTAGCG	CGTGCAAAAG	AGTTGGGCTT	3780
GGATGTGAAA	AGCCACTCAT	CAAGTGTTGA	AGAAGCTGTC	GCTGCAAAAA	TTGCTGCCAG	3840
CTTTAAGCCT	GCAGCTGCTC	CGAAAGTAGA	AGCAAAACCT	GCAGCCCCAA	AAGTAAGTGC	3900
AGAAAAGAAA	GCCGAAAAAT	CTGAGCCAGC	TAAACCAGCT	GTAGCTAAGG	AAGAGGCAAA	3960
ACCTGCAGCC	CCAAAAGCAA	GTGCAGAAAA	GAAAGCCGAA	AAGTCTGAAC	CAGTAAAACC	4020
AGCTGTAGCC	AAGGAAGAGG	CAAAACCAGC	TGAGCCAGTC	ACTCCGAAAA	CAGAAAAAGT	4080
AGCGGCTAAA	CCGCAAAGTC	GTAATTTCAA	GGCTGAGCGT	GAAGCACGTG	CTAAAGAGCA	4140
GGCAGAGCGA	CGCAAGCAAA	ATAAGGGCAA	TAACCGTGAC	CAACAACAAA	ACGGAAACCG	4200
TCAGAAAAAC	GACGGCCGTA	ATGGTGGAAA	ACAAGGTCAA	AGCAACCGCG	ACAATCGTCG	4260
CTTTAATGAC	CAAGCTAAGA	AGCAGCAAGG	TCAGCAAAAA	CGTAGAAATG	AGCGCCGTCA	4320
GCAAGAGGAT	AAACGTTCAA	ATCAAGCGGC	TCCACGTATT	GACTTTAAAG	CCCGTGCAGC	4380
AGCCCTAAAA	GCAGAGCAAA	ATGCAGAGTA	CGCTCGTTCA	AGTGAGGAAC	GCTTCAAGCA	4440
GTATCAGGCT	GCTAAAGAAG	CCTTGGCTCA	AGTAACAAA	CGCAAGGAAC	CAGAGGAAAT	4500
CTTTGAAGAA	GCGGCTAAGT	TAGCTGAACA	AGCACAGCAA	GTTCAAGCAG	TGGTTGAAGT	4560
CGTCCCTGAG	AAAAAAGAAC	CTGCAGTGA	TACACGTCGT	AAAAACAAG	CTCGACCAGA	4620
CAAAAATCGT	GACGATTATG	ATCATGAAGA	AGATGGTCCT	AGAAAAACAC	AAAAGAATCG	4680
AAGTAGTCAA	AATCAAGTGA	GAAATCAAAA	GAATAGTAAC	TGGAATAACA	ACAAAAAGAA	4740
CAAAAAAGGC	AATAACAAGA	ACAACCGTAA	TCAGACTCCA	AAACCTGTTA	CGGAGCGTAA	4800
ATTCCATGAA	TTGCCAACAG	AATTTGAATA	TACAGATGGT	ATGACCGTTG	CGGAAATCGC	4860

1203

AAAACGTATC AAACGTGAAC CAGCTGAAAT TGTTAAGAAA CTTTTCATGA TGGGTGTCAT	4920
GGCCACACAA AACCAATCCT TGGATGGGGA AACAAATTGAA CTCCTCATGG TGGATTACGG	4980
TATCGAAGCC AAACAAAAGG TTGAAGTGGA TAATGCTGAC ATCGAACGTT TCTTTGTCGA	5040
AGATGGTTAT CTCAATGAAG ATGAATTGGT TGAGCGTCCA CCAGTTGTTA CTATCATGGG	5100
ACACGTTGAC CACGGTAAAA CAACCCTTTT GGATACTCTT CGTAACTCAC GTGTTGCGAC	5160
AGGTGAAGCA GGTGGTATTA CTCAGCATAT CGGTGCCTAC CAAATCGTGG AAAATGGTAA	5220
GAAGATTACC TTCCTTGATA CACCAGGACA CGCGGCCTTT ACATCAATGC GTGCGCGTGG	5280
TGCTTCTGTT ACCGATATTA CGATCTTGGT CGTAGCGGCA GATGACGGGG TTATGCCTCA	5340
GACTATTGAA GCCATCAACC ACTCAAAAGC AGCTAACGTT CCAATCATCG TAGCTATTAA	5400
CAAGATTGAT AAACCAGGTG CTAACCCAGA ACGCGTTATC GGTGAATTGG CAGAGCATGG	5460
TGTGATGTCA ACTGCTTGGG GTGGAGATTC TGAATTTGTT GAAATTTCCG CTAAATTCAA	5520
CCAAAATATC GAAGAATTGT TGGAAACAGT CCTTCTTGTG GCTGAAATCC AAGAACTCAA	5580
AGCAGACCCA ACAGTTCGTG CGATCGGTAC GGTATCGAA GCGCGCTTGG ATAAAGGAAA	5640
AGGTGCGGTC GCAACCCTTC TTGTACAACA AGGTACCTTG AATGTTCAAG ACCCAATCGT	5700
TGTCGGAAAT ACCTTCGGTC GTGTCCGTGC TATGACCAAC GACCTTGGTC GTCGTGTTAA	5760
AGTTGCTGGA CCATCAACAC CAGTCTCTAT CACAGGTTTG AACGAAGCAC CGATGGCGGG	5820
TGACCACTTT GCCGTTTACG AGGATGAAAA ATCTGCGCGT GCAGCAGGTG AAGAGCGTGC	5880
CAAACGTGCC CTCATGAAAC AACGTCAAGC TACCCAACGT GTTAGCCTTG AAAACCTCTT	5940
TGATACCTTT AAAGCTGGGG AACTCAAATC TGTTAATGTT ATCATCAAGG CTGATGTACA	6000
AGGTTCTGTT GAAGCCCTTT CTGCCTCACT TCAAAAGATT GACGTGGAAG GTGTCAAAGT	6060
GACTATCGTC CACTCAGCGG TCGGTGCTAT CAACGAATCA GACGTGACCC TTGCCGAAGC	6120
TTCAAATGCC TTTATCGTTG GTTTCACAGT ACGCCCTACA CCACAAGCTC GTCAACAAGC	6180
AGAAGCTGAC GATGTGGAAG TCCGTCTTCA CAGCATTATC TACAAGGTTA TCGAAGAGAT	6240
GGAAGAAGCT ATGAAAGGGA TGCTTGATCC AGAATTTGAA GAAAAAGTTA TTGGTGAAGC	6300
GGTTATCCGT GAAACCTTCA AGGTGTCTAA AGTGGGAACT ATCGGTGGAT TTATGGTTAT	6360
CAACGGTAAG GTTGCCCGTG ACTCTAAAGT CCGTGTTATC CGTGATGGTG TCGTTATCTA	6420
TGATGGTGAA CTCGCAAGCT TGAAACACTA TAAAGACGAC GTGAAAGAAG TGACAAACGG	6480
TCGTGAAGGT GGATTGATGA TCGACGGCTA CAATGATATT AAGATGGATG ATGTGATTGA	6540
GGCGTATGTC ATGGAAGAAA TCAAGAGATA AGATTTTTTG CTCCTTTCTT AGGTGGTGAG	6600

1204

GGACGCAAGC	AAACCGATGG	TTTCATTGCT	TATTTTGTAG	CCTAGGGTCT	CAAAAATCCC	6660
CTGTGATGGG	ACTGATAAAT	CAGTTCCATC	ACTTTCACCA	CGGCGAAAGA	AGCAGATGAC	6720
TTCAAATTGA	ACTTCGTTTC	AATTTAAACT	GAAAATCAAG	AAGTTTAAAA	TAGCTAGGTC	6780
TGCTGGCCTA	GCTTTTGGTT	CAAAGTAGAG	AAAGGAATAT	CATGGCAAAT	CATTTCCGTA	6840
CAGATCGTGT	GGGCATGGAA	ATCAAGCGTG	AAGTCAATGA	GATTTTGCAA	AAGAAAGTCC	6900
GTGATCCACG	TGTCCAAGGT	GTGACCATCA	TAGATGTTCA	GATGCTGGGT	GACTTGCTCTG	6960
TTGCCAAGGT	TTATTACACC	ATTTTGAGTA	ACCTTGCTTC	GGATAACCAA	AAAGCCCAAA	7020
TCGGGCTTGA	AAAAGCAACT	GGTACCATCA	AACGTGAACT	TGGTCGCAAT	TTGAAATTGT	7080
ACAAAATCCC	AGATTTGACC	TTCGTCAAAG	ACGAGTCCAT	CGAGTATGGA	AACAAGATTG	7140
ACGAGATGCT	ACGCAATCTG	GATAAGAACT	AAAGAAGAGG	GGTTGCCCTT	CTTTTTTGGT	7200
GGAGGAAAAT	AGGTTGAATT	TGAAATGGAA	AAATATTCTT	TTATAATAGA	TTGAAACTAG	7260
AATAGTACGC	CTCTACTTCT	AAAATATTGT	TAGAAATCGA	TTTGAATGTC	CTGATCGATT	7320
TGTCCTGTTC	TTGTTTCATT	TTAATATAAA	AAAGGGATTC	TGTATTTTTT	AATGTTATCT	7380
AATTAGAAAA	TGCTTTTTTT	GTAGGAAATA	TAATATGATA	AGGTGCAAAA	AAGAAATAAG	7440
GAGTTTGTAT	ATGGCTGAAC	AAGACTTAGC	TATGCAAGTA	TTGCAACAAG	TGGTGAAACT	7500
ACCTGTTGTT	AAGGTTGATC	GTTGCAAAAT	TTTAGTGGAT	AAGTTTTCCA	AAGAATTGGA	7560
TCCAAAAGAT	ATTCCTACCT	TATTGGAACA	AGGTCCAACG	ACTCTTCTAT	CTCAAGAAAT	7620
ATTAGATCGT	GTAGCTAATG	CTTGATTTG	GGACAATGTA	TTATTAGCGA	GTGGGACTTC	7680
TGTTTTGGCA	GGATTACCTG	GAGGGCTTGC	TATGGCAATT	ACCATTCCAG	CTGATGTGGC	7740
TCAATTTTAT	GCTTTCTCTC	TGAAATTGGC	TCAAGAATTA	GGTTATATTT	ATGGTTATGA	7800
GGATCTTTGG	GCTTCACGAG	AGGAGTTGAG	TGAAGATGCT	CAAAAATACC	TCTTGCTTTA	7860
TCTAGGCGTA	ATGTTAGGGG	TGAATGGAAC	CGCTGCTTTG	CTACGTGTTG	GTAAGTATAAC	7920
AATTGCCAAA	CAGGTAATGA	AAATAGTGCC	TAATAAAGCT	TTAACAAGA	CGCTTTGGTA	7980
CCCTATTTTG	AAAAAAGTCT	TAAAAATATT	TGGTGTGAAT	CTTACCAAGG	GAGGGTTGGC	8040
CAAAGGAATG	GGGAAATTTA	TTCTATCTT	GGGTGGTATC	ATTCAGGTG	GTTTAACCTT	8100
TGCAACTATG	AAACCAATGG	GGGAAAGCTT	GCAGAAAGAA	TTATCCAAGC	TAGTCAACTA	8160
TAGTGAAGTT	CAATATCAAG	AAGATGTTGA	AACAATCCGA	AAAGAGGCTG	AAATCATCAA	8220
AGGAGAGTAA	TATGAATCCT	ATCAAAGCTT	TTGCTAAAAT	TTATGGTAAT	TACTTTTTGA	8280
CCGTGCAAGG	TGTAAGAGTG	ATGAAAACGA	TAAAGAAAGC	TGACCATGTC	GTTGTTGGTC	8340
TGGGGAAACT	TTTTATTGCC	GACAAGTTAA	TGGATACGGC	TCGGTGGCTC	ATTAAGCCAG	8400

1205

AGGAGAGAGA ATGAAATTTT TTTGGTCTTC TTGCTATTCT TTTTATCAAA CCGATTATTG	8460
GGATTGTGAA ATTCTTTTGG ATGATCATCT CTTTTCAGT CCAATGCTG TTTTACAAGA	8520
TAGTGTTTAA GATATTGGAT TGGCTCTTTA AACTTATCTA GATGGTAATC CAAGTTGCAG	8580
AGAACTAGCA GGAACCTCCAC TGCTAGTTTT TTATTCTCTT TCCATATGGT ATAATATAAG	8640
CAGTAAATC ATTTTATACT CTTGAAAAAT CTCTTCAAAC CACGTCAGCT TCACCTTGCA	8700
GTATATATGT TACTGACTTC GTCAGTTCTA TCCACAACCT CAAAACGGTG TTTTGAGCTG	8760
ACTTCGTCAG TTCTATCTAC AACCTCAAAA CACTGTTTTG AGCAACCTGC GGCTAGCTTC	8820
CTAGTTTGCT CTTTGATTTT CATTGAGTAT TAGAACATAC AATGGAGGTC GTCATGGACA	8880
ATATCATCGA TGTGTCAATT CCTGTTGCAG AAGTGGTGA CAAGCATCCA GAAGCTTGG	8940
AAATCTAGT GGAGTTGGGT TTAAACCCC TTGCCAATCC CTTAATGCCG AATACAGTTG	9000
GTCGTAAAGT ATCACTTAAA CAGGGTTCTA AGCTAGCAGG AACTCCTATG GACAAGATTG	9060
TACGCACACT GGAAGCGAAT GGCTACGAAG TGATTGGATT AGACTAATGA CAGATGAACG	9120
GATTCATATC CTACGGGATA TTTGTTAGA ATTGCACAA TGGCGCTCTC CTGAGTCGGT	9180
TCAAGATCGC TTTGATGCGA CCTTTACGGG CGTGTACGCC ATCGAGATTT CCCTTATGGA	9240
GCACGAGCTG ATGAACTCGG ATTCGGGCGT CACTTTTGAA GATGTTATGG AACTCTGTGA	9300
TGTCCATGCC AATCTTTTAA AAAATGCTAT CAAAGGTGTC GAAGTTTCAG ATACTGAGCA	9360
TCCAGGTCAC CCAGTTCGTG TCTTCAAAGA AGAAAATCTG GCTCTCCGTG CGGCCTTGAT	9420
TCGCATTCGT AGATTGTTAG ATACCTATGA GTCTATGGAA GACGAGGAAA TGCTGGCGGA	9480
GATGCCGAAG GGTGTTGGTC GTCAGATGGG ACTTGTGGGT CAATTTGACA TCCATTACCA	9540
ACGTAAGGAA GAACTCTTCT TTCCTATCAT GGAGCGCTAT GGACACGATT CACCTCCCAA	9600
AGTTATGTGG GGAGTGGATG ATCAGATTAG GGAACCTCTT CAAACAGCTC TAACGACAGC	9660
CAAGTCACTA CCAGAAGTGT CAATTAGCAG TGTAAGGAA GCTTTTGAAG CTTTTCGAC	9720
AGAGTTTGAA AGTATGATTT TCAAGGAAGA GTCCATCCTC CTCATGATTC TCCTTGAGTC	9780
TTTTACTCAG GATGACTGGC TTCAGATTGC GGAGGAGAGC GATGCCTATG GCTATGCCAT	9840
CATCCGTCCG TCAGAGAAAT GGGTGCCAGA ACGACAGAGC TTTATTGAGG AAAAGATTGC	9900
AGAGGAGCCT GTACAGCTAG ATACGGCAGA AGGTCAAGTT CAACAAGTCA TAGATACGCC	9960
AGAAGGCCAT TTTACCATTA CCTTTACCCC TAAGGAAAAG GAAGCTGTGC TGGACCGCCA	10020
TAGTCAACAG GCTTTTGGTA ATGGCTATCT TTCAGTCGAG CAGGCCAATC TCATCCTCAA	10080
TCATCTCCCT ATGGAGATTA CCTTTGTCAA TAAAGAAGAT ATTTTCCAGT ATTACAATGA	10140

1206

CAATACGCCA GCTGATGAGA TGATTTTCAA	ACGGACGCCG TCCCAAGTCG GGCGCAATGT	10200
CGAACTCTGC CATCCGCCTA AGTACTTGGA	CAAGGTCAAA ACTATCATGA AGGGGCTTCG	10260
TGAGGGAAGC AAAGACAAGT ATGAAATGTG	GTTCAAGTCT GAGTCGCGAG GTAAGTTTGT	10320
CCACATCACC TATGCTGCAG TACACGATGA	AGACGGAGAA TTCCAAGGAG TGTTCGAGTA	10380
TGTTTCAGGAT ATCCAGCCCT ACCGTGAGAT	TGATACGGAC TATTTTCGTG GATTAGAATA	10440
AGGAGAAAAA ATGAGTTACG AACAAGAATT	TATGAAGGAA TTTGAAGCTT GGTCAATAC	10500
CCAAATCATG ATTAACGACA TGGCGCACAA	GGAAAGCCAA AAAGTTTACG AAGAAGACCA	10560
GGACGAGCGT GCCAAAGATG CCATGATTCT	CTACGAGAGT CGCTTGGATG CTTATCAGTT	10620
CTTGCTTGGT AAGTTTGAAA ACTTCAAAGT	AGGCAAGGGA TTCCATGATT TGCCAGAAGG	10680
CTTGTTTGGT GAGCGAAATT ATTAACGAG	AAAGATTCTT GATTTTTCAC TAAATCTTG	10740
ATAGAATGTT TATGTTAAAT CCTTGTGAGA	GCAGGGATTT TTTATTGAAA GGATTTTATC	10800
ATGTCAAAGA AACTCAATCG TAAAAACAA	TTACGAAATG GCCTCCGTCG CGCAGGTGCC	10860
TTTTCAAGTA CGGTGACTAA GGTGTAGAT	GAGACAAAAA AAGTCGTGAA GCGTCAGAA	10920
CAGTCAGCAA GCGCAGCTGG TAAGGCTGTT	TCTAAAAAAG TTGAACAAGC AGTAGAAGCT	10980
ACCAAAGAGC AAGCTCAAAA AGTAGCTAAT	TCTGTAGAAG ATTTTGCAGC AAATTTGGGT	11040
GGACTTCCAC TTGATCGTGC CAAGACTTTC	TATGATGAAG GAATCAAGTC TGCTTCAGAT	11100
TTCAAAAACT GGACTGAAAA AGAACTCCTT	GCCTTGAAAG GAATCGGCCC AGCTACCATC	11160
AAGAAATTGA AAGAAATGG CATCAAGTTC	AAGTAATTTT TCTTGAGCCT TGCATTTCGG	11220
AAAAATCTT GCTACAATAG AGCCATTAGA	GGTGTMTTGA ATCCACATT TTACAGAAAG	11280
TGGCGGCGCT GAGAAGTCCA CAAATGTGTC	AAAAGTGGT GCTAATGGAT GAAAAATTGA	11340
AATAAAAGTG TCTTTTGTCT TTAAAGACGA	GAGTTGCG	11378

## (2) INFORMATION FOR SEQ ID NO: 211:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4156 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

CCGCGAGCCA CGGCGAATTT GCTGCGGGTA	TTCATCAGTC AGGATCTATG ATCTTTGGTG	60
AACAAGAAAA GGTTCAGTT GTGACCTTTA	TGCCAAATGA AGGTCCTGAT GATCTATACG	120
CTAAGTTTAA TAACGCTGTT GCTGCATTG	ACGCAGAAGA TGAGGTTCTA GTTTTGCGTG	180

1207

ACCTTTGGAG TGGTTCTCCA TTAAACCAAG CTAGTCGCGT GATGGGAGAA AATCCTGAGC	240
GTAAGTTTGC CATCATCACA GGACTIONACT TACCAGATGT GATTCAAGCC TACACAGAGC	300
GCCTCATGGA CGCTGCTGCA GGTGTAGAAA AAGTCGCTGC TAATATCATT AAAGAAGCCA	360
AAGATGGCAT CAAAGCTCTT CCAGAAGAGC TAAATCCAGT CGAAGAAGTT GCAAGCGCTG	420
CAGCTGCTCC AGTTGCCCAA ACTGCTATCC CAGAAGGAAC TGTATCGGA GACGGTAAAT	480
TGAAAATCAA TCTTGCCCGT CTTGACACAC GTCTACTTCA CGGTCAGGT GCAACTGCTT	540
GGACTCCAGA TTCAAAGCA AATCGTATCA TCGTTGCTTC AGATAACGTG GCTAAAGACG	600
ACCTTCGTAA AGAATTGATT AAACAAGCAG CTCCAGGTAA TGTCAAGGCT AACGTGGTTC	660
CAATTCAAAA ACTGATTGAG ATTTCAAAAG ACCCACGTTT TGGAGAAACA CATGCCCTTA	720
TCTGTGTTGA AACACCTCAA GATGCCCTTC GTGCCATCGA AGGCGGCGTG CCAATCAAGA	780
CTCTTAATGT TGGTTCTATG GCTCACTCAA CAGGTAAAAC ATTGGTCAAT ACCGTTTGT	840
CTATGGACAA AGAAGACGTT GCTACATTTG AAAAAATGCG TGACTTGGGT GTTGAATTTG	900
ATGTCCGTAA AGTACCAAAT GATTCTAAAA AAGATTTGTT TGACTTGATT AACAAAGCCA	960
ATGTCAAATA AGCCATTATT TATGAAAGGA TTTTAAACAT GTCTATTATT TCTATGGTTT	1020
TAGTAGTCGT TGTAGCCTTC TTTGCAGGTC TTGAAGGCAT CCTCGACCAG TTCCAATTTT	1080
ACCAACCACT TGTAGCCTGT ACCCTTATTG GGCTTGTAAC AGGTCACTTG GAAGCAGGGA	1140
TTATCCTCGG TGGATCGCTT CAAATGATTG CCCTTGCTTG GTCAAATATC GGTGCTGCTA	1200
TCGCTCCTGA TGCTGCACTT GCTTCTGTCG CTGCTGCCAT TATCATGGTT CTTGGTGGTG	1260
ACTTTACCAA GACTGGTATC GGTGTGCCC AAGCGGTTGC TATCCCTCTT GCTGTAGCTG	1320
GACTTTTCTT GACAATGATT GTTCGTACAA TTTCAGTTGG TTTGGTTTCACT ACTGCAGATG	1380
CTGCCGCTAA AAAAGGTGAC TTCGGCGCTG TGGAGCGTGC GCATTTTCATC GCGCTACTTT	1440
TCCAAGGACT TCGTATCGCG CTTCTGTCAG CTCTTCTCCT TATGGTACCA ACTGAAACTG	1500
TACAAAGTAT CCTTAGTGCC ATGCCAGACT GGCTCAAAGA TGGTATGGCT ATCGGTGGTG	1560
GTATGGTCGT TGCCGTTGGT TACGCCATGG TTATCAACAT GATGGCAACT CGTGAAGTAT	1620
GGCCATTCTT CGCTCTTGGT TTCGTTCTCG CTGCTGTGTC AGATATTACT CTAATCGGAT	1680
TCGGTGCTAT CGGCGTTGCT ATCGCTCTTA TCTACCTTCA CCTTTCTAAA ACTGGTGGA	1740
ATGGTGGCGG AGGAGCCGCA ACTTCTAACG ACCCAATCGG CGATATCCTA GAAGACTACT	1800
AAGATAAGAA AGGACTGAAA ACATCATGAC TGAAAAACTT CAATTAACATA AATCAGATCG	1860
TAAAAAAGTT TGGTGGCGTT CAACCTTCTT ACAAGGTCT TGGAACTTTG AACGGATGCA	1920

1208

AAACTTGGGC	TGGGCTTATA	CACTCATTCC	AGCTATCAAA	AAACTCTATA	CTAAAAAAGA	1980
AGATCAAATC	GCTGCTCTTG	AGCGTCACCT	TGAGTTCTTC	AACACTCATC	CATACGTAGC	2040
TGCTCCAGTC	ATGGGGGTTA	CTCTTGCCT	TGAAGAAGAA	CGTGCTAACG	GTGTGGAAAT	2100
CGATGACGCT	GCTATCCAAG	GGGTAAAAAT	CGGTATGATG	GGACCTCTTG	CTGGTATCGG	2160
TGACCCAGTA	TTCTGGTTTA	CAGTACGCCC	AATCCTTGGA	TCTCTCGGTG	CTTCACTTGC	2220
CCTTACTGGC	AATATCTTGG	GGCCACTCCT	CTTCTTTGTT	GCATGGAAC	TGATTTCGTAT	2280
GTCATTCTTG	TGGTATGTTT	AAGAGATTGG	ATACAAGGCT	GGATCAGAAA	TCACTAAAGA	2340
TATGTCTGGT	GGTATCCTTC	AAGATATCAC	TAAAGGAGCT	TCTATCCTTG	GGATGTTTAT	2400
TCTTGCTGTC	CTTGTTCAAC	GCTGGGTAAA	TATTAAATTT	GCTTTCGATG	TTTCTAAAGT	2460
TCAACTAGAT	GAAAAGGCTT	ATATCCATTG	GGATAAATTG	CCAGAAGGGT	CTAAAGGTAT	2520
CCAAGAAGCA	TTCCGACAAG	TAGGACAAGG	ATTGCTCTCA	ACTCCTGAAA	AAGTTACTAC	2580
TTTCCAACAA	AACTTGGATA	TGTTGATTCC	TGGATTATCA	GGACTACTCC	TTACTTTACT	2640
TTGCATGTAC	TTACTTAAGA	AAAAAGTATC	TCCAATCACT	ATTATCCTTG	CCCTCTTCGC	2700
AGTGGGTATT	GTGGCACATG	TTCTTCACAT	CATGTAATCA	AGCAACTAAA	AAGGAACCAG	2760
GTTCTAAAAT	CTGATTCCTT	TTTTCTATGC	TTTTATTTCAG	CCAAGGCTCC	CATTGGATCC	2820
CATGGTGCAA	GTACGATTGG	TTCTGCTCCA	TAGGCAGCTT	GTTCTTCTGC	TGTCAGCAAT	2880
TCCTTACGAA	CAACGATTTG	GTATGTGTAT	TCGTCATCC	AAGCGTCTGA	GGCAACAAAG	2940
TAACCATCTG	TACCGACCTT	GTCTCCCCAT	GAGTTTTCAA	CCTTCCACTT	GGTTGATTTA	3000
CCATTTTCGT	CCAAGTCAAC	ACCTGTCAAG	ACCATGGCGT	GGGTCACTAA	GCTTTCCTTA	3060
TAGTCCAAAC	GTCCAGCCTT	GTCTTGAGTA	AGTTTAATGT	CCATGCTTGA	TTCAAAGTCA	3120
TAAACATCTG	TCGCAAGGAT	GCCAGCTTAC	GGTTGCTGAG	CTGGCCGACA	TCAGAACCAA	3180
ACCAAACAGT	CTCACCTGCT	TGCATTTGGG	CAATCGCCAA	TTCTTTCAAG	CGCTCCATTG	3240
GAACGTTGAT	GTAGCGAACT	GCACGGCTAC	CAACCACATT	CCCCAACATC	TCAACTGTGT	3300
AAGATTTTCC	GTAAGGTTTA	TCAGCAGTTG	GAGCATTGAT	AACAGAAACG	TAGTCTTCTA	3360
AAGGAAGATT	GACATATTTT	TTGTAAAACT	CTTGTTGGTG	GATTCCTTTT	TCACTTTTGT	3420
AGTTGTTATC	TTTATCGCGA	TAAGCAAAGT	CAAACTTGCG	TGGTGGAAAG	CCTAATGACA	3480
TAGCAAGAAA	GTTAAAGATT	TCTTGCAAGA	GGTCTTCTTT	CTTAGCTTGA	ACAGTCGCTT	3540
GATCTGCACC	AGAAACAAGC	AAGTCACGCA	AGATTGAGC	ATCTTGACGA	AGCAATTTAT	3600
TAAGGATCGC	ATTAGCTCA	CGACTGCTGC	TAGATGAAAC	AGACTCAGGA	TAAACTGACT	3660
TAGGCACGAC	ACCGTATTTT	TCAAAGAGGG	AAACGACCAT	ATCCCATTGA	CCGCCATCTT	3720



1209

GTTGAGGTGT TTGGAGTAAG AAGCTAACtT GCGGCTAGTC AATTCTTGCT CTGAAGTCGC	3780
AATGACTTGC TCCAAGAACC AGTTTGATTT CTCATACTTA TCCCAGAAGA AAGTGTGGGC	3840
TTGTGACAAC TCAAAGTTCT CCAATTTGTA TTGCGAGATG AGTTTGTGGC GGAAGGTGTT	3900
GAGAGCCGCA AACATCCAGC AACGACCAGA CGCTTTCTGG TTAGTGACCT TGTCCTTGGT	3960
TAAATCCAAT GAGAAAACAG GTGTGTTGTC TACATGGCTT TGGCGACGTT CCAGAGCTGC	4020
AAAAATTCCG TTGTGGCTGG CAGCATTTTC AATCGCTTGG TATTTTACAT TTGCTTCATA	4080
GTTGGCAAAT AGTTTATCAG TAAATGATTC TTGAATCGCG TTCATAGATT CCTCCTTTTA	4140
GTCTACAGTG TATTGG	4156

(2) INFORMATION FOR SEQ ID NO: 212:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3902 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

AAAAACAACA AAATAAAACA AAAACAAAAA TATCGAGGTT TATTTTCAAA ACTTTCGATA	60
TTTTTATTAA GTTATTATTT TGTTGTTTCT AGTTTACTTT TTGATGGTTA AGAGTGGTGG	120
AGAATTATAC TCAATGAAAA TCAAAGAGCA AACTAGGAAG CTAGCCGCAG GCTGTACTTG	180
AGTAGCGCAA GGCGAAGCTG ACGTGGTTTG AATTTGATTT TCGAAGAGTA TTAGTGCAAA	240
CCGTAGTTGT AGTCATCATC TTGCATGGCT TCAACTTCGC CAAGAAGGTA ACCATTTCGG	300
ACTTGAGAGA AGAAGTCATG GTTGGAAGTT CCTGTTGAAA TACCGTTCAT AACGATTGGG	360
TTGACATCTT CAGCTGAATC TGGGAAAAGT GGATCTTGTC CCATGTTCAT GAGAGCTTTA	420
TTGGCATTGT AGCGAAGGAA GGTTTAAACC TCTTCAGTCC AACCAACACC GTCATAAAGA	480
CTCTCTGTGT AGCCTTCTTC ATTTTCATAA AGAGTATAGA GTAGGTCGTA CATCCATTCT	540
TTGAGTTTTT CTGCTCTTC TTCAGGTAAT TCATTGAAAC CAAGTTGGAA TTTGTAACCA	600
ATGTAGGTTT CGTGAACAGA CTCGTCACGA ATAATCAATT TAATGATTTT TGCAACGTTG	660
GCAAGTTTGT TGTTACCGAG ATAGTAGAGG GGAGTGAAGA AACCAGAGTA GAAGAGGAAG	720
GTTTCGAGGA AGACGCTGGC AACTTTCTTT TCAAGTGGGC TGCCGTTTAG GTAGATTTCTG	780
TTGACAATCT CAGCCTTCTT TTGTAGGTAA GGATTGGTAT TGGTCCATTG GAAAATTTCT	840
TCAATCTCAG CCTTAGTATT CAAGGTAGAA AAGATTGATG AGTAAGATTT AGCGTGGACA	900

1210

GATTCCATAA ATTGGATGTT ATTGAAGACA GCTTCCTCAT GTGGTGTACG GATGCTGCG	960
CGAAGGGCTT GAACCCAGT TTCAGATTGC ATAGTGTCAG GAAGGGTTAA ACCACCAAAA	1020
ACTTTTCCGA CCAAGTCTTT CTCTTTGTTA GATAGCTTTC TCCAGTCATC CAAGTCGTTT	1080
GATAAGGGAA TACGTGTATC GAGCCAAAAT TGCTCCGTCA GTTTTCCCA AGTTGATTTG	1140
TCGATGACAT CTTGATGGC ATTCCAGTTA ATGGCTTTGT AGTAAGTTTC CATTTAAAAAT	1200
CTCTTTCTGT GTTTAGTATT GCGAACTCAC AATTATTTCT ACTTTACCAT AATTCTATAG	1260
GAGTATCGCA CAAAAGTCG GAAGCCCGAC TTTAAAAATG TTACATAAAT TATGTTATGA	1320
CATAGTAGAT TTGATTTTAT CAGTGCTGCT TAGGGAAAAA TAGTGTTTCT ATGCTAGAAA	1380
CTAAATCACA CAGCTTTCAC ATTGGTTGGC GCCGACTTCT CCACCGTCAT CTGTAAAGGT	1440
ACGGACGTAG TAGATAGACT TGATTCCCTT GTTAAAGGCA TAGTTACGAA GGATGGACAA	1500
GTCACGTGTC GTTTGTTTAT TTTCCCTCTT CCATTCGTAA AGGCCTTTTG GAATGTCACT	1560
GCGCATGAAG AGGGTGAGTG AAAGTCCTTG ATCCACGTGT TCAGTCGCAG CAGCGTAAAC	1620
ATCGATGACT TTACGCATAT CCATATCGTA GGCAGAAGTG TAGTAAGGAA TGGTTTCTGT	1680
AGACAAGCCA GCAGCAGGGT AATAGATTTT ACCAATTTTC TTCTCTTGGC GTTCTTCGAT	1740
ACGTTGCGTA ATCGGGTGA TAGAAGCAGA AACGTCGTTG ATATAGCTGA TAGAACCATT	1800
TGGCGCTACA GCAAGGCGAT TTTGGTGGTA AAGACCATCT TCTTGAACCT TGTCGCGAAG	1860
TTCAGCCCAA TCAGCAACAC CAGGGATAAA GACATTTTGT AAGAGTTCTT TAACACGGTC	1920
TGATGTTGGA ACAAATTCAC CAGTTACATA CTGTCAAAG TAACTTCCGT TAGCATAGTC	1980
TGATTTTCA AAGTTGTGA AGGTAATACC ACGTTCACGT GCAATATTGT TTGACTCTAC	2040
CAAGGTCCAG TAGTTCATAA GCATAAAGTA GATGCTTGTA AATTCAACAG ACTCAGGTGA	2100
ACCATATTCA ATGAGTTGTT GGGCAAGGTA GCTGTGCAGT CCCATGGCAC CGAGACCAAA	2160
GGTGTGGGCT TGGCTATTTT CATGGTCAAT CGTTGGTACA GCTACGATAT GTGAACATATC	2220
TGTAACGAAA GTAAGGGCAC GAACCATAGC ACGGATAGAA CGACCAAAAT CAGGTGAAGT	2280
CATCATGTTA ACCACGTTGG TTGAACCCAG GTTACATGAA ACATCTGTTT CCATTTGAAG	2340
GAATCTTGA GCATCGTTGA TCAAGCTTGG TTCTTGAAC TGAAGAATCT CAGAACACAA	2400
GTTACTCATG ATAATCTTTC CATCAACAGG ATTTGCACGG TTAGCCGTAT CGATGTTGAC	2460
TACATAAGGA TAGCCAGACT CTTGTTGCAA TTTAGAGATT TCAGTTTCCA AATCCCAGGC	2520
CTTGATTTTT GTCTGCGAA TATTTGGATT TGCGACCAAT TCATCGTATT TTTAGTAAT	2580
GTCGATGTAA TTGAATGGCA CACCGTATTC TTTTCTACA GAGTAAGGGC TGAAGAGGTA	2640
CATTTCTTCA TTTTACGAG CCAATTCGTA GAATTTATCA GGTACTACAA CACCAAGTGA	2700

1211

TAGAGTCTTG ACACGTACTT TTTCATCAGC GTTTTCTTTC TTAGTTGAAA GGAAAGCGAT	2760
GATATCTGGG TGAAAGACGT TGAGGTAGAC AACACCAGCA CCTTGACGTT GCCCCAATTG	2820
GTTGGAGTAA GAGAAGCTGT CTTCAAAAAG CTTCATAACA GGAACGACAC CTGAAGCAGC	2880
TCCTTCATAG CCTTGATAG GTGCACCAGC TTCACGAAGG TTGCTGAGGG TAATTCCCAC	2940
ACCACCACCA ATACGTGAAA GTTGAAGAGC TGAGTTGATA GAACGCCCGA TAGAGTTCAT	3000
ATCATCCGTC ACTTGATTG GGAACAAGA TACCAACTCC CCACGACGAG CACGTCCAGC	3060
ATTCAAGAAG GAAGGAGTAG CAGGTTGGTA GCGTTGGTGG ATGATTTTCAT TGGCAATATC	3120
GATTGCAACA GCTTCATTCC CATCAGCGAA ATAAAGGGCA TTGAAGAAGA CACGGTCTTC	3180
CATATTTTCA AGATAGTATT CACCGTCATT AGTCTTTAAG GCATATTGAT TGTAAATTT	3240
ATAAGCTGCC ATGAATGACT TGAATTGGAA GTTTTGGTCT TTGATAAATT GAGCTAATTC	3300
TTCCAAGAAC TCTGGACGGT ATTTCTTGAT AAAGGCTGTT TCGATGTAGT TGTGTTCAAT	3360
GAGGTAATTG ATTTTGTCTT TGATTGAATC AAAAACCATA GTGTTTGGA CTACATTTTC	3420
TTTAAAGAAA GCATCCAAGG CTTCTTGTC TTTATGAAGC ATGATTTGTC CATTACAGG	3480
ACGGTTAATT TCGTTATTAA GACGGAAGTA AGTCACGTCT TCAAGATGTT TTAATCCCAT	3540
AAAATTTCCC TTATCTAATT ACAAAGAAA GGCTTCTAAG TTAGCCCTAA AAGCAGTTTC	3600
TTCTGGATGA TGTACTAAGA TTATGCTAAT TGTTCAGTT TTCCTGGTTG GAAACCTGAA	3660
AAGACTTCAG TTGGTGTGTT GATAACAGGA GCTGCGCTAA AACCAGAGCTC TTTAACTTGA	3720
TCGACGTACT CAGGTTGCTC ATCAAGATTG ATTTACAGAT AAGAGACATT ATTACTGTCC	3780
AAGAAACGCT TGGTCATTTT ACATTGGACA CAATTGTTTT TAGAATAAAC GGTACCATT	3840
GTGTAACCTC TCTTCAAAT TTAATACTAT CTTAGTATAT CAGAAAATAA AATTTGTGCG	3900
GG	3902

## (2) INFORMATION FOR SEQ ID NO: 213:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2456 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

TATTGAAGCT ATTGTAGACT ACAAAGATAA GGATTGTCAG TTAGTAGGCG GTGAGACTCA	60
CTGATAACCT AAAAAGGATA GTCAATTATG CTTGTTTACT AACTATTAAC TATGCTAAAT	120

1212

CAATTGAGGT TGTTCACATA AAACCTCTATA TCAGAGAAGC CTGATATAGA GTTTTTCTT	180
GCTAGTTTGA GGATTTTTTT GTAAAATAGA AAAAGTGAAG AGAGGTATGA AATGAGCAAG	240
AAAGATAAAA AAATCGAAAT TCAAGTAGCG GATGCCAAAG TTAATGTTGC TAAAGACAGT	300
TTTGAAGGTT ATACATTGAC TATCGGTAAA AAAGTTATCG GAGAAATTGC CGAATTAGAC	360
GGACAATTTG CCATTATAAA GAATGGGAAT GTCGATAGTT TTTATAAAAA ATTGGAAAAA	420
GCTGTGGAAG TTTTGATTGA AAATTATAAT TTAGCAAAAT AAGTCTTGTT TTTTGAAT	480
TTTCATGATA TAATAGTCCA TGTGATTGT AGGAGAGATA GCGAAGAGGC TAAACGCGGC	540
GGACTGTAAA TCCGCCCTT CGGGTTCGGG GGTTCGAATC CCTCTCTCTC CATTTTCATTA	600
ATGGGGTATA GCCAAGCGGT AAGGCAAGGG ACTTTGACTC CCTCATGCGT TGGTTCGAAT	660
CCAGCTACCC CAGTTCTTAG GTAATAATCA AGATAGAAAG CAAAATATCT TAGGGTATTT	720
TATTTTATA ATTGAAAGAC GTGAATGATA TGAACATGTC CTGCGGGTG CTTAGGAAAA	780
AAATTATAAG TATGTCAAGT TTAAGAAAAA CTGATTGTT GGAGGATTTT TTAGATGAAC	840
GAATTTGAAG ATTTGCTAAA TAGCGTTAGT CAAGTTGAGA CTGGTGATGT TGTAGTGCT	900
GAAGTATTGA CAGTTGATGC GACTCAAGCT AACGTTGCAA TCTCTGGAAC TGGTGTGAA	960
GGTGTCTTGA CTCTTCGCGA ATTGACAAAC GATCGTGATG CAGATATCAA TGACTTTGTT	1020
AAAGTAGGAG AAGTATTGGA TGTCTTGTA CTTCGTCAAG TAGTTGGTAA AGATACTGAT	1080
ACAGTTACAT ACCTTGATC TAAAAACGC CTTGAAGCTC GCAAAGCATG GGACAAACTT	1140
GTTGGTCGCG AAGAAGAAGT TGTTACTGTT AAAGGAACGC GTGCCGTAA AGGTGGACTT	1200
TCAGTAGAAT TTGAAGGTGT TCGTGGATTT ATCCCAGCTT CAATGTTGGA TACTCGTTTC	1260
GTACGTAACG CTGAGCGTTT TGTAAGTCAA GAATTTGATA CTAAATCAA AGAAGTTAAC	1320
GCTAAAGAAA ACCGCTTCAT CCTTTCACGT CGTGAAGTTG TTGAAGCAGC TACTGCAGCA	1380
GCTCGCGCTG AAGTATTCGG TAAATTGGCT GTTGGTGATG TTGTAAGTGG TAAAGTTGCT	1440
CGTATCACAA GCTTCGGCGC TTTCTGCGAC CTTGGTGGTG TTGACGGATT GGTTCACTTG	1500
ACTGAATTGT CACATGAACG TAATGTATCA CCAAAATCAG TTGTAAGTGT TGGTGAAGAA	1560
ATTGAAGTGA AAATCCTTGA TCTTAACGAA GAAGAAGGAC GTGTATCACT TTCACTTAAA	1620
GCAACAGTAC CAGGACCATG GGATGGCGTT GAGCAAAAAT TGGCTAAAGG TGATGTAGTA	1680
GAAGGAACAG TTAAACGTTT GACTGACTTC GGTGCATTTG TTGAAGTATT GCCAGGTATC	1740
GATGGACTTG TTCACGTATC ACAAATTTCA CACAAACGGA TTGAAAATCC AAAAGAAGCT	1800
CTTAAAGTTG GTCAAGAAGT TCAAGTTAAA GTTCTTGAAG TTAACGCAGA TGCAGAACGC	1860
GTGTCACTTT CTATTAAAGC TCTTGAAGAA CGTCCAGCCC AAGAAGAAGG ACAAAAAGAA	1920

1213

GAAAAACGTG CTGCTCGTCC ACGTCGTCCA AGACGTCAAG AAAAGCGTGA TTTCGAACTT	1980
CCAGAAACAC AAACAGGATT TTCAATGGCT GATTTGTTTG GTGATATCGA ACTTTAATCA	2040
AATTGAAAAT TCACAAAATC CTTGTTTAC TAAACAAGGG ATTTTCTGG CTCTTTGTCA	2100
ACTGTAGTGG GTTGAAGAAA AGCTAAGCTC GAGAAAGGAC AAATTTTGTC CTTTCTTTTT	2160
TGATATTCAG AGCGATAAAA ATCCGTTTTT TGAAGTTTTC AAAGTTCCGA AAACCAAAGG	2220
CATTGCGCTT GATAAGTTTG ATGAGATTAT TGGTCGCTTC CAGTTTGGCG TTAGAATAGT	2280
GTAGTGAAG GGTGTTGACA AGCTTTTCTT TATCTTTGAG GAAGGTTTAA AAGACAGTCT	2340
GAAAAATAGG ATGAACCTGC TTAAGATTGT CCTCAATAAG TCCGAAAAAT TTCTCCGGTT	2400
CCTTATTCTG AAAGTGAAAC AGCAAGAGTT GATAGAGCTG ATAGTGGTGT TTCAGG	2456

(2) INFORMATION FOR SEQ ID NO: 214:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10974 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

AAATAGGATA TAGAGACATC CTTCTGATCT GCTTTTWACA AAGTCCAATT ATATGCGGAT	60
CTATACCTCC ACAATGTCCA TTATTATmCC TAACTATAAT ATGAGCCGAA AACACTATAT	120
CCTTAATGTC TCCATATCCA TCAGGGATAT TAATATTAT TTTCCACAA CTATATTGCA	180
TTGTAACCAT CTCCTTAAAC GACGCATTAT GATATTTGAT AGAGAAATTT TTATGAATAA	240
CTCAATAATT TTATAGTAAA TCATGCTTAT ATCTCAAAGA TACCTATTTT ATCTTGCTC	300
GACCTTCTCC AAAGAATTGC TATAATACTA TTACAAATCC ATCTGCACTA CACTTCAAAT	360
TTTAGCACTG TATAAAAACG TTTCAATACA CTAACCTCAA GAAAACCTCC ACTATTAATT	420
GAAAAAATTG ATAGAGATAA ATTAAAAATC TATATTGAAA CTCATCCCGA TGCTTATTTG	480
ACTGAAATAG CTGCTGAATT CAACTGTCCT CCAACAAC TAATTACGC TCTAAAGGCT	540
ATGGGATATA GTCTAAAAAA GAGCCGTACC TACTGCGAAC AAGACCCAGA AAAAGTAAAT	600
CGGTTCCCTTA AAGAATTGAA TCACCTAAGC TACCTGACTC CTATTTATAT TTATGAGACA	660
GGGGTTGAGA CCTATTTTTA TCTCGAATAT GATCGAGCCT TGAGCAGGCA GTTAGTCTCT	720
CTGGAAGAAG ATATAATTAT TTGAATTAAG ATCGAGACAA CGCACACCAG AGATTGCGAT	780
ACTGTTATAG AAGTACTAAT GCCCTTTTTT GTTCAATAT ACTATGGCTC CGATGACCTA	840

1214

TAAAGATACG ATGACGAGTG ACTTTTTCGA AGCTTGCTTC CAAAAATTCT TACTACCTAC	900
TTTAGATACA CCATCCCTTA TCATTATGGA CAATGCAAGG TTTCACAGAA TGAACATGTG	960
TAAGGAGCAG GGCATAGACT GTTACCACTT CCTACCTATT CACCCGAGTA TAATCCCATT	1020
GAGAAAATAT GGGCTTACAT CAAAAACATC TCAGAATAAT ATTGTCAAAT TACGATGCTT	1080
TTCTTGAGGC ACTTTTGTCC TATTCTTGTT TCAGCCGACT ATACTCCGTT ATTGGGCAGC	1140
TACGGAACAG TCGATGGGAC GATGGGGGGA CATAAAAAAA TCCTCCAGTT TTGTTTTTTA	1200
TAACAGTATA CTGGAGAATT GACAATCTCG GTAGATACCT CGTTATAGCG CGGTTACTTA	1260
TTAGGCAGTT AAAAAACAAC TGTGAACAGA AAACATTCCA GAGTCAGACA AGACTTTGGA	1320
ATGTTTTGGC TCTATAATTT CTGTAGTGGG TAATCCCACC CCAGGAATTA TAGGGTCGTT	1380
TCTTGTAGAA AAAAAGCCCC ATATGACCTA TAATGAAAAG CGTCTAACCA ACTCATTAGA	1440
AAGGGTTCAT ATGGAACAAC TTAAGAATAC CACAGATTTG CTCGGATTGG AAGACAAAAA	1500
TATCAAAATC TTGTCTGTTC TGAAATACCA AACCCTCTA GTCGTTCAGG CAAAGTTGGA	1560
TTCCCCCGCT CCTCCTTGTC CTCATTGTCA AGGGAAGATG ATCAAATACG ACTTCCAGAA	1620
AGCCTCTAAA ATTCCGCTTC TCGACTGTCA GGGTTTACCC ACGGTACTGC ATCTCAAAAA	1680
GCGCCGCTTT CAGTGCAAGA ATTGCCTTAA GGTGGTCGTT TCTCAAACAT CCATTGTCAA	1740
GAAAAATTGC CAGATTTCCA ACATGGTGAG ACAAAAAATC GCTCAGCTCC TCCTTGAAAA	1800
GCAGTCTATG ACTGAGATTG CCCACAGATT GGCGGTCTCA ACTTCCACCG TCATCCGAAA	1860
ACTGAGGGAA TTAAAGTTTG AAACCGATTG GACCAAGTTG CCAAAAGTTA TGAGTTGGGA	1920
TGAGTATAGC TTCAAAAAGA GCAAAATGAG CTTCAATTGCC CAAGATTTTG AGTCCAAATC	1980
CATCCTCGCA ATTTTAGACG GGCGAACTCA TGCGGTGATT CGAAACCATT TCCAACGCTA	2040
TCAGAGAGAG GTTCGGGAGC TGGTCGAGGT CATCACCATG GACATGTACA GCCCTTATTA	2100
TGGGCTCGCT AAGCAACTCT TTCCAAAGGC GAAGATTGTT CTGACCGCT TCCACATTGT	2160
CCAACATCTG AGCCGAGCTA TGAACCGAGT ACGAATCCAA ATCATGAACC AATTTGACCG	2220
AAAATCCTTG GAGTATCGGG CGCTCAAGCG CTTTTGGAAC CCTCGCTTTT TCGTTTCTAG	2280
GCTCGGGCTA AATCAGTCCA CTGGACTGAT TTA CTACACC AGTATAGCTT CAAGCTCTGT	2340
CAGAAACGAT TCTATCAGCC CACGTTTCGA ATGCACTTAA CCCATCGGGA AGTACGAGAT	2400
AAGCTGCTTT CTACTCTGA GGGATTACAG GTTCACTACG AACTCTATCA ACTCCTGCTC	2460
TTTCATTTTC AAGAGAAGAA TGCCGACCAT TTCTTTGGAT TGATTGAGCA AGAACTGCCA	2520
ACGGTTCATC CGCTTTTTC AACGGTCTTT TGGACTTTT TAAGGGATAG AGATAAGATT	2580
ATCAACGCAC TTAAGCTGCC TTATTCCAAC GCTAAACTTG AAGCGACCAA TAATTTGATT	2640

1215

AAGATTATCA AGCGCAAAGC CTTTGGTTTC CGGAACCTTA ACAATTTTAA AAAACGGATT	2700
TTGATGACTT TGAACATCAA AAAAGAGAGT ACGAATTTTCG TACTCTCCAG ATTGCAGCTT	2760
TTGCGCTACC CACTACACTT GACAAAGAGC CACTCTTTAT TCCATGGTAT CAAAGGCAAG	2820
ACTTGGTTTG GCATTGAGGT CCCAGCCTGC GAAGTTTCTT TTGTTCCACT CGCTGACGCT	2880
GGCATAGGCA ATCATACTTG CATGTCTCC GCAGAGTCGC AGAGGGGGGA TGATAACCTT	2940
GACATCTGTG ATTTTCGGCTG CTAGGCGTTC TCTGAGACCT TTATTGGCTG CCACACCACC	3000
TGCCACAACCT AGGATTTTAA CAGGATATTT CTCCAAAGCC TTCTTGGTMT TTGCCATGAG	3060
AATGTCCATA ACTGCTGCTT GGAAGGAAGC ACACAAATCT TCTGTAGACA GGCTTTCTCC	3120
CTTTTGCTCG GCATTGTGAT GAAGATTGAT AAAGGCAGAT TTCAAACCTG AGAAGGAGAA	3180
CTCCAGATTA TCTTCCTTAA TCATGGCAGC GGGGAAATCA TAAATATCCT GCCCCTGATG	3240
AGCCAGCTCG TCAATCTCAC GACCTGCAGG ATAGGTCAAG CCCATGACAC GGCCGACCTT	3300
ATCATAAGCC TCACCAACCG CATCATCAGC GGTTCCTCCA ACAATCTTAT AATCTCTGCT	3360
CTCCGAAACA TAAACCAACT CTGTGTGTCC GCCGCTGACC AAGAGGGGTA GCAAGGGAAA	3420
CTCCAAAGGC TCCACACTCT GAGCTGCCAT GAGGTGCCCA GCCATGTGAT TAACAGGAAT	3480
CAGTGGAAGT CCGTGAGCCC AAGCAAAGGC CTGGCAGCT GACAAACCAA CTAGCAAGGC	3540
TCCGACCAAG CCTGGTCCGT AGGTAACCGC AACAGCTGTC ACGTCCTCTT CGGTAATCCC	3600
TGCTTCTGCC AATGCCTCCT CGATACAGGC TGTAAATGACC TCGACATGGT GACGACTGGC	3660
TACTTCGGGC ACTACGCCAC CAAAACGTTT GTGACTCTCA ATTTGACTAG CAATGACATT	3720
GGACAAGAGC TCATCGTCGT TTTTCAAGAC GGCGACACTG GTCTCATCAC AGGATGTCTC	3780
AAATGCTAAA ATATATCTAT CCTTCATCTA TTTCTCTCTT CATGATAATG GCGTCCTCGA	3840
CTGGGTCAATG GTAGTAGGCC TTTCGCTCAG CGATAACTGT CATCTTTTCT TTCTTGTAAG	3900
ATGCTTGCGC TCGTTGATTT GACTGTCTGA CTTCGAGGAA AATTTCTTTG TCTGTGGCA	3960
ATTGAGCAAA CAAGGCTGAC GCAATCCCCT GACCCTGATA AGCTCCTTTG ACAGCGATTT	4020
GCAGGACTTC TGCTTCAAAA AGATTCTCCT GCACAGCTAG AAATCCAATC ACTTCTGCCC	4080
CATCATAAGC CAATGCATAC CAAGTCTGGT CTGCGGACAG ATCTGCTTGG ATTTGCTCCA	4140
GAGTCCAAGG ACTGACTAGG TAAACAGCTG CCATAACAGC GTAGATGGCT TGAGCTAGGT	4200
CAGGCTGTTG TTGAATTGCG TTGATTTCTA TCATAGGCGT TTAATGTAAG ACTCGCCAGA	4260
CTCGGTATGG TTCTTGAGCC AGTTTCTCTC AGCCTCGACT CGTTTGAGGT AATTCGGCAC	4320
AAAATCATGC AAGGAGTCTG CTTCTTGTG CCAGGCCAAA AGAGCTAGAT TAGCTGCATT	4380

1216

GGGCAATGTT TCTTTGTAAT CAGTCCTTGG CAAGTGTTTT TGAATCTGCT CAACAAAGGG	4440
GCCAACTTCT CCGACAAAGG TTACCTGACT AGTACCCTTG ACTTTTCTA GCACCTCTTC	4500
AAAAGATAGG TGGCCTTCTG CCATGACAGG TTTGGCATTTCATATAAAATC CTGCATAAAC	4560
ATTATTGCGA CGCGCATCCA TCAAGGGGAC AAACAAACCT TCTTGTGTGAT GGGGCACCAG	4620
AGCCAAGAGA CTCGACATAC CAACCAACTC GATGTTTCAGG GTGTGAGCTA AGGTCTTAGC	4680
AGTTGCTACC GCAATTGCGA AGCCTGTATA GCTACCCGGC CCTTCAGCTA CCACGATTTCG	4740
GTCCAAATCC TTGGGTGTCC AATCCAACT TGCCATCAAA AAATCGATGG CAGGCATAAG	4800
AGTAATACTG TGATTTTCT TAATATTAAT CGTCGTCTCG GCAAGAACCT GCTTATCTC	4860
TAAATAGCC AGAGAAAGAG CCTTGCTGGA CGTATCAAAA GCTAATACTT TCATAACACA	4920
TTCTATCTT TTTGTCTGCT TACTATTATA CTACAAAAGC TGGCACATGG GAATTTTCTT	4980
TGCCCCAGA CAAGAGTGCC CTCACTTAAC TAAAAATAAT TTAATAAAT GCTCACTTTT	5040
CCTTTTCTT TCCGAATATA AAAGTGAACA AGAAAAAGG AGGAAAGTTC AATGACAAAT	5100
TTTGACATTC TTGACAATCA ATTTTATCC TTATCTGAAA ATGAATTATC AGATATTGAT	5160
GGCGGTCTCG CTCCTTGGT TATCTTTGGA GTAGCAGTAT CTTGGAAGGC TATTGCAGGT	5220
GGAACAGCAC TTATAGGTTG TGGTTTGGA GCTGGTTATT TTTTAGGAGG AGATTAATAT	5280
GATGAAAGAT TTGAACAATT ATCGTGAAT TTCTAATAAG GAATTGCAAG AAATCAAGGG	5340
TGGCTTTGGT GTCGGTGTG GTATCGCTTT ATTTATGGCA GGTATACCA TTGAAAAGA	5400
CCTTCGTAAA AAGTTTGGA AGTCATGCTA GATAAGAAAC ACATTTTATAG AAGGATAAAT	5460
TTTATTGTCT TCATCTCTTA CAGTTTGCTC AGCATCTCA ATGATTGAA CATTACTACC	5520
ATCCCTTAC CATTCGATTT ATCTGTTGT ATTGTTTAT TTTTATGCTT CAACTCTATT	5580
TTTGATCAGA ACAATGACTC CCATAAAAT AATAAGCTTT GAAAATCCA TTGTCATGTC	5640
ATGTTAGAAA AATGCAAAGA CCACCTCATC TTGATAGATG GGGTGGAATT TTCGTGTCGT	5700
AAATCTACTA TCTCTACATT CCCAAACAAA AAACCCAGC ATAAGCAGGG CATCTAAGCA	5760
TTTAATTCAA AGTAAATAC AAACCAAACG ACATAGGTCA CGAGGAGGAG AAAAGCGAG	5820
TAGAGAGTCA CAAAGGTCAT TTTCCACAAG AACTTGTTTT GTCGTCGTT CAGTTTGGA	5880
AATAGAAGAT TCCCCGCATA AACGCAAGCA AAAAAACAA TAAAAGCTAC CAAGCGAGCT	5940
CCGATAGCAA AAGCAAATAA GTTATACATA GGGCAACCTC CTTGACTTAA AATCTATATG	6000
GAATTATGAC AAGCAATAAA TTTCACTTCC GTTATCAACA TAATACATTT TCTTTATTTT	6060
TGAAAACGCT TACCAAAGAA ATCGTCCCCT AACTTTCTCG TTTCCGTCTT TACTAATTT	6120
TTCAATTTGT GGTATAATTG AAATAATTGT AACGAATCAA GGTCAATCTA GACACAAAAT	6180



1217

GGAATGAAAT CAAGCAAATA TCTGCTAAAA GTTTGGAATA AGCTGACCTG TAAATAGAAA	6240
GGAACTATAT GATTTACAAA GTTTTTTATC AAGAAACAAA AGAACGTAGC CCACGCCGTG	6300
AAACAACACG CACGCTTTAC CTAGACATCG ATGCCAGCTC AGAACTTGAG GGCCGTATCA	6360
CTGCTCGCCA ACTTGTGAA GAAAATCGCC CAGAGTACAA TATCGAGTAT ATCGAACTCT	6420
TGTCTGACAA ATTGCTCGAT TACGAAAAAG AAAGTGGCGC CTTCGAAATT ACGGAGTTCT	6480
AATATGGCCT-ACACTCTTAA ACCTGAAGAA GTGGCGCTTT TTGCCATCGG TGGTCTAGGA	6540
GAAATCGGGA AAAACACTTA CGGAATTGAA TACCAAGACG AGATTATCAT CGTCGATGCT	6600
GGGATTAAAT TCCCAGAAGA TGACTTGCTT GGTATCGACT ATGTCATTCC TGACTACTCT	6660
TACATCGTGG ACAATATCGA CCGCGTCAAG GCTGTTTTAA TCACACACGG ACACGAGGAC	6720
CACATTGGTG GGATTCCGTT CCTACTCAAG CAAGCAATG TCCCTATTTA TGCTGGACCG	6780
CTTGCCCTTG CTTTGATCCG TGGGAACTC GAAGAACACG GCCTCTTGGC CAACGCCAAA	6840
CTTTACGAAA TCAACCACAA CACCGAGTTG ACCTTTAAAA ATCTCAAGGC AACTTTCTTT	6900
AGAACGACTC ACTCTATTCC AGAGCCTTTG GGGATTGTCA TTCATACTCC TCAAGGGA	6960
ATCGTCTGTA CGGGTGACTT TAAGTTCGAC TTTACTCCAG TTGGAGAACC TGCGGACTTG	7020
CATCGTATGG CTGCGCTTGG TGAAGAAGGC GTGCTCTGTC TCCTGTCTGA CTCGACAAAT	7080
GCGGAAGTAC CAACCTTTAC CAACTCTGAA AAAGTCGTTG GTCAGTCCAT TATGAAGATT	7140
ATCCAAGGTA TTGAAGGACG TATCATCTTT GCATCCTTTG CCTCAAATAT CTTCCGTCTC	7200
CAGCAGGCAA CAGAAGCTGC TGTTAAGACT GGACGCAAGA TTGCGGTCTT TGGTCGTTCT	7260
ATGGAAAAGG CCATTGTCAA CGGAATCGAT CTTGGCTACA TCAAAGCTCC TAAGGGAACC	7320
TTTATCGAGC CAAATGAAAT CAAAGATTAT CCTGCAGGAG AAGTTCTTAT CCTCTGTACA	7380
GGTAGTCAGG GTGAGCCTAT GGCAGCCCTC TCTCGTATCG CCAACGGAAC CCACCGTCAA	7440
GTACAATTAC AACCAGGTGA TACCGTTATC TTCTCTTCTA GTCCCATCCC TGGAACACT	7500
ACTAGTGTC AACAAGCTGAT TAACATCATT TCTGAAGCTG GTGTCGAAGT TATCCACGGT	7560
AAAGTGAACA ATATCCATAC ATCTGGACAC GGTGGTCAGC AAGAGCAAAA ACTCATGCTC	7620
TGCTTGATTA AGCCAAAATA CTTTCATGCCT GTCCACGGTG AATACCGCAT GCAAAAAGTC	7680
CACGCTGGAC TAGCAGTGA TACTGGTGTT GAGAAGGACA ATATCTTTAT CATGAGCAAT	7740
GGCGATGTGC TTGCCCTTAC TGCTGACTCA GCTCGTATCG CAGGTCATTT CAACGCCCAA	7800
GATATCTATG TCGATGGAAA TCGTATCGGT GAAATTGGCG CAGCTGTCCT CAAAGATCGT	7860
CGCGATCTAT CTGAAGACGG TGTCGTTCTG GCAGTTGCAA CTGTTGACTT CAAATCGCAG	7920

1218

ATGATTCTAT CTGGTCCAGA CATCCTCAGC CGAGGCTTTG TCTACATGAG AGAGTCTGGC	7980
GACTTGATTG GCCAAAGCCA GCGTATCCTC TTCAATGCCA TTCGTATCGC ACTGAAAAAT	8040
AAGGATGCTA GCGTGCAATC TGTCAATGGT GCCATTGTCA ACGCTATTCTG CCCCTTCCTC	8100
TATGAAAATA CCGAACGTGA ACCGATCATC ATCCCGATGA TCCTCACACC AGATGAAGAA	8160
TAAAGCAAGA AAACAGCCCC GTCCTCGGAG CTGTTTTTCT CTATGCTTTC TTTTGAGATT	8220
AAAATCATA CTCAATGAAA ATCAAAGAGC AAAGTAGGAA GCTAGCCGTA GGTGCTCAA	8280
AGCACTGCTT TGAGGTTGTA GATAGAAGTG ACGAAGTCAG TAGCCATACC TACGGCAAGG	8340
CGACGTTGAC GCGGTTTGAA GAGATTTTCG AAGAGTATCA ATAAAAATCG AAATCAGACT	8400
AGAAGGCTAA GCGAAAGCAT AACTTGAGTT AGCTCCCATG GTTCGGGAAA CTATGGGAGG	8460
CTGGAGATGA ATCAAAGCCA AGCTTTGAAC TCATTCTGTA GAAGCCGACG ACGTATCATT	8520
TTGATTTTGT AAGAGTTTGA GAAATACTAC GATTTTACC TTCCAGATAC ACCATCAAAA	8580
TAGAAATATC TGCTGGGTTT ACTCCCGAAA TACGGCTGGC TTGGCCGATG GTTCTCGGAT	8640
TGATGAGTTT GAACTTCTGA CGGGCTTCGG TTGCGATAGA ATCAATGTCA TCCCAGTCGA	8700
TATTGGCCCG AATGCGTTT TCTTCCATGC GTTTCATCTT GGCAACCTGG TCCATGGCTT	8760
TGGAAATATA GCCTTCATAC TTGATTTCTG TTTCAATCAA TTCGATAATC TTGTCATCCA	8820
AGTCTTCTGC AGCTGGTCCG ATGAAGGCCA CCACATCTTG GTAAGAAACT TCTGGACGGC	8880
GAAGGAATTC CTTGGCTGTC ACTGCATCGG TCAAGGGTTT GAAGCCCATC TCCTCAACCT	8940
TGGCATTGGT TTCCTTGACT GGCTTGAGTT TGATACTGTC TAGGCGCTTC ATCTCATTAT	9000
CAAATTGATT TTTCTTGATT TCAAAACGAG CCCAGCGTTC ATCGTCCACA AGGCCAATCT	9060
CGCGTCCCAT CTCAGTCAAG CGCATATCAG CATTGTCATG ACGAAGAATG AGACGGTATT	9120
CAGCACGACT GGTCAAGAGA CGGTAGGGTT CAATGGTTCC CTTGGTCACC AAGTCGTGCA	9180
TCATCACCCC GATATAACCA TCACTGCGCT TCAAAATCAA TTCAGGCTTG CCTTGGATTT	9240
TCAGAGCCGC ATTGATACCC GCGATAATCC CTTGGCCTGC TGCCTCTTCG TAACCTGATG	9300
TTCCATTTGT CTGACCAGCA GTGAAGAGAC CTGAGATTTT CTTGGTTTCC AAAGTCGCAC	9360
GCAACTGATG AGGCAAGACC ATATCATACT CAATAGCATA ACCTGTCCGC ATCATCTCTG	9420
CATTTTCCAA ACCTTTGATG GAATGCACCA AGTCACGCTG GACATCCTCA GGCAGACTGG	9480
TTGAAAGTCC TTGCACATAG ACTTCCTCAG TATTGCGCCC TTCTGGCTCA AGGAAGAGTT	9540
GGTGACGTTT CTTGTCCGCA AAGCGCACAA TCTTGTCTTC AATCGACGGA CAGTAACGAG	9600
GCCCCACTCC CTTGACCACA CCTGTAAACA TAGGCGCACG GTGGAGGTTG TTTTGGATAA	9660
TCTCATGACT GGTACCATTG GTATAGGTCA ACCAGCATGG TACTTGGTCC TTGACATAAT	9720

1219

CCTCATCAGC	TGAAGTGTAT	GAGAAATGAT	TAGGCACTTC	GTCTCCTGGC	TGAATTTCTG	9780
TCACATCGTA	ATTGATAGAA	GAAGCCTTGA	CACGTGGAGG	GGTTCCTGTC	TTGAAACGAC	9840
CGATTTTCGAG	ACCCAGTTCC	TTGAGATTGT	CAGCTAGGTT	AATAGAAGCC	AAGCTGTGGT	9900
TAGGACCTGA	TGAGTACTTG	AGGTCTCCGA	TGATAATTTT	CCCACGGAGA	GCAGTCCCTG	9960
TCGTCACAAT	AACAGCCTTA	GCAGCATATT	CTTGATGGGT	GGCTGTACGC	ACACCGACAA	10020
CCTTGCCATC	TTCCACCAAA	ATCTCATCAA	TCATGGTTTG	ACGAAGGGTC	AGATTTTCTT	10080
GGTTTTCAAC	CGTCTTGCGC	ATCTCCTTAG	AGTAAAGTTC	CTTGTCAGCC	TGCGCACGAA	10140
GGGCACGGAC	AGCTGGCCCC	TTCCCTGTGT	TTAGCATCTT	CATCTGGATG	TAAGTCTTGT	10200
CAATGGTTTT	GGCCATCTCG	CCACCGAGGG	CATCGACTTC	ACGCACGACA	ATCCCCTTGG	10260
CAGAACCACC	GATAGAGGGA	TTACAAGGCA	TGAAAGCCAG	CATTTCAATA	TTGATGGTCG	10320
CAAGCAGGAC	CTTACAGCCC	ATACGGCTAG	CGGCCAAGGA	AGCCTCAACC	CCAGCGTGTC	10380
CCGCACCAAT	TACAATAATA	TCGTATTCTT	CAGTAAATG	ATAAGTCATG	TTTCTCTCCT	10440
ATTCTCAAG	ATGAATGTGT	CTTAGTTGGC	CTTCCCAATC	TGGTAGGGCT	GTTTTTAAAA	10500
AGACTGGAAC	TAGCTGGATA	TTCTGGAGCT	TATCCAAGTC	AATCCACTCA	CAGGGCTGCC	10560
TTTTCTCATC	TTCTGCATG	GTCAACGGGG	CATCTTCAAG	CAAATCCACC	AGATAATGAA	10620
ACTCGATATT	GTGATAGGAA	ACGCCGTCCA	CTTCAAAACG	ATTTTCAACC	ACAAAAGCTA	10680
GCTGCCCAGC	TTGAGCTTTG	ACACCCAGTT	CTTCCTTCAC	TTCACGGACT	ACCGCGTCTT	10740
CCGTGCTTTC	ATTGACTTGA	ATCGCACCTC	CAATAGTGTA	ATACTTGCCC	TTGTCTTTGG	10800
TAACTAGAAG	CTTGTGATTT	TGGACAATCA	AGGCTGTAGC	CCGAACACCA	AAAACCGTAT	10860
TGTCTACTTT	TGTCCGAAAG	TCTTGTTGAG	TCATTCTTGT	CCTTTCCCTT	AAACGACACA	10920
AAAACAGTCA	AAACTACAAA	GAAGTGCAGG	ACAAAAAAGC	CTGCAACATC	CAGG	10974

(2) INFORMATION FOR SEQ ID NO: 215:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 987 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

CCCGTTATGA	TTATGGATAG	CGCTTTCAAA	TTTTTAAACT	CCTATCCCAT	CCTTTTATCT	60
ATATAATAAG	TGAAAATATA	ATAACTGTCA	AGTAACTGAA	GTGAATTTTA	TAAAAAATT	120

1220

ACAAGCCAAA TTTGTAAAGT TTACACTAAG CCGCTAGgCA ATCGTCTATC AGAATATCCG	180
TTTATTTGTC AATAATCCGA GAAAATCTTG CAACGCTTAG AAGTCTATAA AAACATATCAA	240
CATTTATATG ACTTGCGAAT AGCAATCCTG CTAAACCTTT CCACACTCTA TCTATACAAT	300
CAAGATAAAA ACATGTGTAA GCAAATCTGC TACACTTTAC TGGAGGACGC CAAGAATAAG	360
AAAAGCTACG ATAGGCTTGC TATCTGCTAT GTCCGTATTG GGATTGTAC AGACGATTCT	420
AAACTTATCC AAAAAGGGTT CTCCCTTCTG GAGCTGACCG AGGAAACTTC TATGCTGTCT	480
CATCTCAAAA AAGAAGTAGA GACCCATTAT CAACCAAAGA AATTATAAAA AAAGTCGAGG	540
GAGCTCCTCG ACCTTTTCAT AGAATCGCCG AACGATTTAA CGAGAAAGTA TGACTTTTAC	600
GTTTATCCCA ACTCAATTAT GACATTTTTT TCAAAAGTCA ATATATCTCA CTTTTTCAAC	660
GACAAGAAAG AGGCTGATAA TCTACCAACC TCTTATTCTG AACCCATCAC TCCATCACTT	720
TTTAGCTTCA TTCGCTTTCT TAGCGACTGC AATCTGGTAT TCGACTTGGT CATTTCCCTT	780
ACCGGTACAA CCATGAGCAA TTGTAGTCGC TCCTATCTGA TCGCTATTT CAACCAATTT	840
TTTAGAAATC AGAGGGCGGC TCAAGGCAGA TACCAAGAGA TACTTTTGTT CATAATAGGC	900
ATGTGACTGA TGAGCCACTA GCACATAATC TGTAGCAAAT TCGTCCTTAA CATCAATGAC	960
ATAAGATTCT ACTGCCAAA CCTTAAG	987

## (2) INFORMATION FOR SEQ ID NO: 216:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2651 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

CTGGGTCCTG TTCATAGTAG GTGTGGTtCT TTTTTTCGAG TGAGCCCAT AGCTTTGAGC	60
GCATAGTGGG TGGTAGTTGG ATGACAGCCA AAGTCAGAAG CTATTTCAGT CAAATAAGCA	120
TCTGGATTGT CAGTAAGATA GTTTTAAAGT CTATCTCTAT CAACTTTTCT TGGTTTGTGTT	180
CCTTTTACTT GGTGGTTTAG CTCTCCTGTT TTCTCTTTTA GCTTTAACCA GCCATAAATG	240
GTATTACGTG AGATTTGGAA AACGTGTGAT GCTTCTGTTA TACTACCTAT TCGCTCACAA	300
TAAGAGAGAA CTTTTTTACG AAAATCTATT GAATATGCCA TAAGAAGATT ATACCACATT	360
GTGTACTATT TTTGGTTCAT TTTACTATAT TTTATAAGTT ATAGTGTAGC ATTCCAACCTT	420
CAAAGCACTA TAAAGTAAAT TGAACAAGA ACAATACAAA CAATTCTCGT AAACGGATTG	480
CAACCACAAA AAAGCAAGCA TTCACAAGAA TACTTACCTA TCATGGGAGG AACCAACGTT	540

1221

CCTCTTTTTT	ATTACTAAAA	TTCAAAGAAT	TCCAATGCTT	TTTTCAAGAG	CAAATCCGTA	600
TATTCTGGAT	CTTCTGGGGC	TACTTCTATT	TCCCGCTGAA	CTTTTCCAA	ATCATCTGTA	660
ATCACTCCAT	CTACTCCTAA	GTGAAGAGAT	TTGCTGATAG	CTTCTGAATC	ATTGACAGTC	720
CAGACATAAA	GTTTCTGATC	CGTTGTCCAT	AGTTTGCTTA	CAAAATATTC	ATCCAAGGTT	780
GAGTACTCCA	TAGTATATCC	TGTCGCTCTT	GTTTTAGGAA	AGACAGAATT	GTAGGGCATG	840
ATGAAATAAA	CTGGTAGTTC	GGCATCATAC	TGTCTTACTT	TTTCGACAAC	ATGGTAGTCT	900
AAAGACTGGA	TTTGATGTCC	ATAAATCTTG	AGCTTTGCAG	CATAACGGGC	TAAAAAGCGG	960
TTCATCATGT	CTGGACTATC	TTTTTTACTG	GTTTTAATTT	CAATTAGTAA	TTTTTGACCA	1020
AGTTCGTTGG	CTCGACTGAG	ATAATCTTCA	AAGCTTGAAA	TTTTAGTCTG	GTAGCCATTT	1080
TCAAAAATAT	CAATCCCTTT	AAGCTCCTCC	AAGTTTAAGT	CTTGAGGACT	TTTATTGATA	1140
CCTGCTAGAT	TTTTCAAGTT	AGCATCATGC	ATCATGACAA	ACTGCCCATC	TTTTGTTTCC	1200
TGCACGTCGG	TCTCCACCAA	GTCTGGTTTG	AGTTGTGCTG	TAGTTTCCAA	GGACTCTACT	1260
GTATTTTGAA	TCCCATTTCG	ATTGGAAACC	CCTCGGTGAG	AAATAAGTTG	AGGTAGATGA	1320
ACCATGGGAG	CCTCCAGATA	AATATAACCT	TCTAAGGCAA	AGAAAAGACT	GGCACAAGTC	1380
ATGACACCCC	ATCGCACCAT	GTGATCTTTT	TCTCTCCTAG	GAAGCATATC	CAGCTCCTTT	1440
CCTGTCAAAA	ATGAAACAAA	TTTAACCAAA	AAATAAGTCA	GAGCCATATA	ATAGAGATTT	1500
TTAATCACGA	CAAAATTCAA	AATACCAAGA	ATCAGAGACT	CTCTCTGAGT	GATATCATCT	1560
ACCAAAGTTT	GAGCCAATAA	TAAAGGAATC	AAAGGAAGAT	AATAATAATA	ATGTGCTTTG	1620
AGCAAGATGT	AAAATAAATT	CCAAGCATAA	AAAGTAACTC	TCTTCTTGGT	TTTCTCCAAG	1680
CTAAACATCA	CTGCTTCTCG	AACAGTCAGC	TGATCATATA	CAATCTTCGG	AAGGGCAAAC	1740
ATCAATCTGA	CAGAGACATA	GAGAAAGATA	AGAGATAGAA	GTAGGATGCT	CAGCCACCAC	1800
ATCCAATATC	TATCTTCTAA	ATAAGCTTGG	ATAAACTCTG	GAATGACGAT	TTTATTAAGA	1860
TAATAAATCT	TCAGCATTTT	CCGTATAAAA	GGAACAGCA	TAGCTATATA	GAAAAAGATA	1920
AACAAGGCTT	TAGCGCAAGT	TAGCTTTTTC	ATAAATCCAA	AACCTTCATG	GAAAACCTTG	1980
CGGATATACT	CAATTAGCCT	TCGCTTTTCA	TTATAGAGGA	GATGACGAGC	ACCAATAAAG	2040
AGGAGTCCTA	TTTGAAAATA	AGCAACCAGA	AGGTTAATTA	CAATCAAGGC	TAAAAAAGCT	2100
AGACTAATCA	ATGGAGAATG	AGTAAGGATG	GCTAAGACAT	TGTTATAGGA	AATAAAAAGA	2160
TAACCTGTCT	GATCTAATAA	GAAGCTAGCC	AACCATGAAT	TGAATGGTAC	CCACAAATAC	2220
TCCACTATCA	TAAAAATCAA	GAAAAATAGA	AAGAGGATTT	TATCAAGATC	GAGGTAATC	2280

1222

TGTTTAAGAC CCAATTTTTT AGGTTTTTCA GGTTCATAG GCACTCCTAG TCAAATAATT	2340
GAGACAAGTC CAAGCCACCA AAAGGATTGT TTGATAAGCT ACTTTCTGTC TCTAACAATT	2400
CCCTAGCTTG ATCCGACTCT AAGAAGGATT CGTAAACACG CGCCGTCATC CGAGCATCCT	2460
CTAAACTATT ATGAGACTGA CCTTGAAATC CAAGAAATGA GGCAACAGTT TGCAATTTGA	2520
GATTGGCAAT ACCATGTAAA TCTGAACTCC GACGTTCAAA AGCTTCATCA TACAAATCCA	2580
CCTTGTA CTG TTGGCTATAG TCTAAACCAT GCTCTGCTAA AATAGGTAAA TCACTTTTAG	2640
CAGCATTGTA G	2651

(2) INFORMATION FOR SEQ ID NO: 217:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5638 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

CGTTATAATA AACTTGTGAA AAAATTAACA AAGGATATCG TTCCTTGAAA GCTATGGAGG	60
AAAATATGGC TGATAAAAAA ACTGTGACAC CAGAGGAAAA GAAACTCGTT GCTGAAAAAC	120
ACGTAGATGA GTTGGTTCAA AAAGCTCTAG TTGCCCTTGA AGAAATGCGT AAATTGGATC	180
AAGAACAAGT TGA CTACATC GTTGCCAAAG CATCAGTAGC AGCTTTGGAT GCCCACGGAG	240
AATTGGCTTT ACATGCCTTT GAAGAAACAG GACGTGGTGT ATTTGAAGAC AAAGCAACTA	300
AGAACTTGTT TGCCTGTGAA CACGTAGTAA ACAACATGCG CCACACTAAG ACAGTTGGCG	360
TTATCGAAGA AGACGATGTA ACAGGATTGA CTCTTATTGC TGAACCAGTT GGTGTTGTTT	420
GTGGTATTAC TCCAACAACA AACCCAACAT CAACAGCAAT CTTCAAATCA TTGATTTTCA	480
TGAAGACACG TAACCCAATC GTCTTTGCCT TCCATCCATC AGCACAAGAA TCATCTGCTC	540
ATGCAGCTCG TATCGTCCGC GATGCAGCTA TCGCAGCTGG TGCTCCTGAA AACTGTGTGC	600
AATGGATTAC TCAACCATCT ATGGAAGCAA CAAGTGCCTT TATGAACCAC GAAGGTGTTG	660
CGACAATCCT TGCAACAGGT GGTAAATGCCA TGGTTAAGGC GGCTTATTCA TGTGGTAAAC	720
CAGCTCTTGG GGTAGGTGCC GGAAACGTT CAGCTTATGT TGAAAAATCA GCAAACATTC	780
GTCAAGCAGC ACACGATATC GTCATGTCTA AATCATTTGA TAACGGTATG GTCTGTGCAT	840
CTGAACAAGC AGTTATCATT GATAAAGAAA TTTACGATGA ATTTGTAGCA GAGTTCAAAT	900
CTTACCACAC TTACTTTGTA AACAAAAAAG AAAAAGCTCT TCTGAAGAG TTCTGCTTCG	960
GGCTCAAAGC AAACAGCAAA AACTGTGCTG GTGCAAAATT GAACGCTGAC ATCGTTGGTA	1020

1223

AACCAGCAAC TTGGATTGCA GAACAAGCAG GATTTACAGT TCCAGAAGGA ACAAACATTC	1080
TTGCTGCAGA ATGTAAAGAA GTTGGCGAAA ATGAGCCATT GACTCGTGAA AAATTGTCAC	1140
CAGTTATTGC AGTTTTGAAA TCTGAAAGCC GTGAAGATGG TATTACTAAG GCTCGTCAAA	1200
TGGTTGAATT TAACGGTCTT GGACACTCAG CAGCTATCCA CACAGCTGAC GAAGAATTGA	1260
CTAAAGAATT TGGTAAAGCT GTTAAAGCTA TTCGTGTTAT CTGTAACCTCA CCTTCTACTT	1320
TTGGTGGTAT CGGGGACGTT TACAATGCCT TCTTGCCATC ATTGACACTT GGATGTGGTT	1380
CTTACGGACG CAACTCAGTT GGGGATAACG TTAGTGCCAT TAACCTCTTG AATATCAAAA	1440
AAGTCGGAAG ACGGAGAAAT AACATGCAAT GGATGAACT TCCTTCAAAA ACATACTTTG	1500
AACGTGATTC AATTCAATAC CTTCAAAAAT GTCGTGACGT TGAACGTGTC ATGATCGTTA	1560
CTGACCATGC CATGGTAGAG CTTGGTTTCC TTGATCGTAT CATCGAACAA CTGGACCTTC	1620
GTGCGAATAA GGTGTGTTAC CAAATCTTTG CGGATGTAGA ACCGGATCCA GATATCACAA	1680
CTGTAAACCG TGGTACTGAG ATTATGCGTG CCTTCAAACC AGATACCATC ATCGCACTCG	1740
GTGGTGGGTC TCCAATGGAT GCTGCCAAAG TAATGTGGCT CTTCTACGAG CAACCAGAAG	1800
TGGACTCCG TGACCTTGTC CAAAAATTCA TGGATATCCG TAAACGTGCC TTCAAGTTCC	1860
CATTGCTTGG TAAGAAGACT AAATTCATCG CGATTCCAAC TACATCTGGT ACAGGATCTG	1920
AAGTAACACC ATTTGCCGTT ATCTCTGATA AAGCAAACAA CCGTAAATAC CCAATCGCTG	1980
ACTACTCATT GACACCAACT GTGGCAATCG TAGATCCTGC TTTGGTATTG ACAGTTCCAG	2040
GATTTGTTGC TGCTGATACT GGTATGGACG TATTGACTCA CGCGACAGAA GCATACGTAT	2100
CACAAATGGC TAGTGACTAC ACTGATGGTT TAGCACTTCA AGCCATTAAA TTGGTCTTTG	2160
AAAATCTCGA AAGCTCAGTT AAGAATGCAG ACTTCCACTC ACGTGAGAAA ATGCATAACG	2220
CTTCAACAAT CGCTGGTATG GCCTTTGCCA ATGCCTTCCT AGGTATTCTT CACTCAATGG	2280
CCCATAGAT TGGTGCGCAA TTCCACACAA TCCACGGTCG TACAAATGCT ATCTTGCTTC	2340
CATACGTTAT CCGTTACAAC GGTACACGTC CAGCTAAGAC AGCAACATGG CCTAAGTACA	2400
ACTACTACCG TGCAGATGAA AAATACCAAG ATATCGCACG CATGCTTGGA CTTCCAGCTT	2460
CTACTCCAGA AGAAGGGGTT GAATCTTACG CAAAAGCTGT CTACGAACTC GGTGAACGTA	2520
TTGGGATCCA AATGAATTTT AGAGACCAAG GAATTGACGA AAAAGAATGG AAAGAACATT	2580
CTCGTAAATT AGCCTTCCTG GCTTATGAAG ACCAATGTTT ACCAGCTAAC CCACGTCTTC	2640
CAATGGTAGA CCATATGCAA GAAATCATCG AAGATGCATA CTATGGCTAC AAAGAAAGAC	2700
CAGGACGCCG TAAATAATTG TTTATCAGTC TAGAAGCAAG AAAAAAATC AATTTGAGGG	2760

1224

AAAGATCCAG TAATTTTCT ATGATAAAAG GCATCCTATC AAGGTTTTTG AACACCTGAT	2820
AGGATGCCTT TTTATGATAT TGAGGCCTTT TTGCCCTTTT TGAAAACTA GAATAGAAAC	2880
AAAAATATATA ATAGATTGAA ACTAGAATAG TACATATCTG CTTCTAAAAC ATTGTTAGAA	2940
TTGATTTGA CTGTCCTGAT CGATTTGTCC TGTCTTATT TCATTTTGAT ATATAAAAAA	3000
TATAGTATAG TAGACTGAAT CTAAAATAGT ACGAAACAAT TGCTAAAACA TTTATAGAAA	3060
TTAATTTTAC TTTTCTGATA GAGTTGTCA CATCTTATTT CAATTCAC TAAGTTTAATT	3120
TAAGAGTAGT ATTTACTAAG GCCCAATTAA AATCAAAGAG CAACTAGAA AACGAGTGCC	3180
ATTCAGCTCA AAACACTGAT TTGAGATTGC AGATAAGACT AGCCCCCTCA TTAACAGATT	3240
TACGATAAAA CGATGACAAG GTGTGTTGCT TTTTGATTTC TAAAGAGTAT AATGATAGAT	3300
CTCTATAAAA TAAGTGCGAA GGAAATGAGC TTTTATAGTC CTTTCGTTTT AAAATACTAT	3360
CTCAGATATT CTTATATCGA CAAGAAGTTT TTGAGTCATT CCCTCATCAT ACATATTAATA	3420
TAAATAGTGG CTCATTCAAT TTTTCACTAG AATAATAAGC TAGTATAGTA AACTGAAATA	3480
AGATATAAAC AAATAAATTG GAGCTTAACA TCCATTCCA GCAATTTTTT AGAACTACA	3540
GTGGACTATT CTAGATTCAA CATATTATAA AAAGTAGAGT AAAAGAAAAG GATTGGATCT	3600
TGTGTAATGC AGGATCCAAT CCTTCAATC ATTTTGTTCA ACTTTTGAG GTTCCTACAA	3660
TGTAGTCGTC ATTAATAAAG ACAGATGGGA ATGACAGTGT TCCTATTTAT TTTGATAGAG	3720
ATCGATGAAT TCTTTAGATA GCAACTGAAT AATCTCTGTT GAAGCCATTT GGTCTTCTGC	3780
ATGCATAAAT AGCAAGGAGA ATCCTATTTT TTCTCCAGTA GCTTCTTTTT GTATGAGATT	3840
AGAGTGAATC TTGTGCGCTT CTAATAAGGA GTCTCCGCT TCTTCAACTT TAATTTTCGC	3900
TTCTTTTAAA TTTCTGCCT TAGCTAGTTG GATGGCTTCA ATAAAGGATG ATTTGGCTGC	3960
TCCACTATTG GCAATGAGCT GAAAACAGAT ATATTCCATT TCTTCTGTCA TCTTATTCT	4020
CCTATCCATG CAAGTGCTTG TTCCAGAACT TTTGCTCCAT TCATCATTCC GTAATCCCGC	4080
ATATCAATGG TATCTACAGG GATATTTCCT GCAATTTCTT TCACAGCAAG TAACTCATAA	4140
CGAATTTGTG GCCCAATTAG AATGACATCT GCTTCATGGA TATTCTTTTT AGCTTCTGTC	4200
ATTGATTTTG CTTGGATAGA GATTCAATC CCACGTTTCA TCGCACTTTG TTGCATTTTT	4260
TTAACAAGCA TACTTGTCGA CATTCCCGCA TTACATACTA ATAAAATTTG TTTCATAATC	4320
TTAACCTTCC ATTTCTTGTT CAACAACCTT GTCATTAATC TTGATAAATG GAATGTATAG	4380
AAGAACTCCA AGTGCAAAGA TGATGAATTG AACTAGAACT GCTCTCACGT CCCCTGCTGT	4440
TGCTAACCAT GCATTTAAGA ATACTGGTGT AGTCCAAGGA ACTTGTATAA ATGCAGGACT	4500
CATGAATTCT GTAAGTGTG CTAAGTAGCT GATTAATAA CCAAGGACTG GAACTGTGAT	4560



1225

AAATGGAATA GCTAATGAAA TGTATAAAC GATTGGGTAA CCGAATAATA CTGGTTCATT	4620
GATATTGAAG ATACCAGGTC CAAAAGATAA TTTAGCCACG TTTTGTAGAGA CAGCATTGCG	4680
ACTCACTAAG AATGTTGCTA TTAATAAACA TAATGTAGAT CCACTACCAC CCATTAAAGC	4740
GAATGTTTGT ATTTGTGATA GGTGATGAT GTGTGGAATG GCTTGTCCAT TATTTGCTGC	4800
AGTGATGTTT TCAGTAATGT TAATTAATAG TAATGGTCT AGGATGGCAC TGTAATAAC	4860
TGCTTGGTGA ATACCAAATA GCCATAACAT ATTTCTAAA GAGTAAATAA TAATGACCCC	4920
GATTAAGCTT GTACCAATAT GACGAATTGG TTCTTGAATA AAGATTGTAA TGATTGAGAT	4980
TAAGTTCATT CCAGTTATAT TGAATAATAA TGCTGAAACA ACCCCAAATA AGGAGATGAC	5040
GGTCATGACT GGAAGTAATA CGCTAAATGA TCTACTAACA GCTGGTGGAA TATTTTCACC	5100
AAGGTTTATT TGTAAGCTT TAACGTTTGA TAATTCAATG AATAATTCTG TTGCAATAAT	5160
CGTACGATAA CCCCAGCGAA CATTGCGCCT GTACCTGTGT TGTGTAATGA AAGAACACCT	5220
GAAATGTTTA CCGCATCTTT TGCTCCGTCA GGAACACAG AACTGTATT TGGCATCATC	5280
ACAATTAAAG AACTAATGA TAGCATTGAT GCTGCTAACG GGTTCGAA ATCTCTGTTT	5340
TTAGCTAAGA AATAACCAAC CATTACAGCA ATAATCATAC CTGAAATACT TAAAGTACCG	5400
TTTGCAATTG TTATTCCTCA ATATTGGAAT CTGTGTAATG TATCCCCTTG GAAAATCCAC	5460
TTAAATACCG TGTGTTCAA AAGAACGATT AAACCTGCCA AAATATATAA TGGCATTACT	5520
GTTACGAATG CATCTCTTAG GTTTTAA TGAATTTGGT TCCCTAGTTT ACCAGCAAAG	5580
GATGGCAAAA AAATTTTTTT GGGGGGGGGG GTTATTAAAC CCCCCTTTT AAAAAAA	5638

(2) INFORMATION FOR SEQ ID NO: 218:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4745 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

CCGGAAGCTG TTGCCCTTGG AACTCCAAAT GAAGAAACAG CCTTGTCTT GAACTATTTT	60
GGTGTGGAAG CACCACGTGT TATCACTTCT GCCAAAGCAG AGGGGGCAGA GCAAGTTATC	120
TTGACTGACC ACAATGAATT CCAACAATCT GTATCAGATA TCGCTGAAGT AGAAGTTTAC	180
GGTGTGTAG ACCACCACCG TGTGGCTAAC TTGAAACTG CAAGCCCACT TTACATGCGT	240
TTGGAGCCAG TTGGATCAGC GTCTTCAATC GTTTACCGTA TGTCAAAGA ACATGGTGTA	300

1226

GCTGTGCCTA AAGAGATTGC AGGTTTGATG CTTTCAGGTT TGATTTGAGA TACCCTTCTT	360
TTGAAATCAC CAACAACACA CCCAACAGAT AAAATCATTG CTCCTGAATT GGCTGAATTG	420
GCTGGTGTGA ACTTGAAGA ATATGGTTTG GCAATGTTGA AAGCTGGTAC CAACTTGGCT	480
AGCAAATCTG CTGAAGAATT GATTGATATC GATGCTAAGA CTTTGAACCT CAACGGAAAT	540
AATGTCCGTG TTGCCCAAGT GAACACAGTT GACATCGCTG AAGTTTGGGA ACGCCAAGCA	600
GAAATTGAAG CTGCAATGCA AGCTGCCAAC GAATCAAACG GCTACTCTGA CTTTGTCTTG	660
ATGATTACAG ATATCGTCAA CTCAACTCA GAAATCTTGG CTCTGGTGC CAATATGGAC	720
AAGGTCGAAG CGGCTTTCAA CTTCAAACCTT GAAACAATC ATGCCTTCCT TGCTGGTGCC	780
GTTTCACGTA AGAAACAAGT GGTACCTCAA TTGACTGAAA GCTTTAATGC GTAAGATTTT	840
GGGTGTCAGC TCAAAATCGG AAAGTCTAGT TTGCCTTATA TCGCAAGGAG TTTGGGCTCC	900
TTTTTTCTAG GAGTGAAGTA TGTTAGAAAA TGGCGATTTG ATTTTGTGA GAGATGGGTC	960
AGACATGGGA CAGGCCATCC AGACTTCCAC AGGTAACAT AGCCATGTTG CCATTTATTT	1020
GGATGGGATG ATTTATCATG CTAGTGGACA GGCTGGTGT GTCTGTCAAG AACCGGCAGA	1080
CTTCTTTGAG TCCAATCATT TATACGACCT CTATGTTTAC CCAGAAATGG ATATCCAGTC	1140
GGTGAAGGAA AGAGCTTGCA AACATCTTGG AGCACCTAC AATGCTTCTT TCTATCCAGA	1200
TGCAGCTGGT TTTTACTGCT CCCAGTATAT AGCAGAAATC CTACCTATTT TTGAACTAT	1260
TCCTATGAAA TTTGGAGWTG GGGAGCAGGA GATTAGTGAT TTTTGGAGGG AGTATTACAT	1320
AGAACTAGGT CTGCCTGTTT CTCTGAACCA AGCTGGTACC AATCCTAGTC AGTTGGCAGC	1380
ATCGCCTCTG TTACAATGTA AAGAAAGGAA TCTTCATGAT TCAGATTTTT AATCCATCTC	1440
GTTTGACGAG ACAGCCATTT TTGGAGAATT GATCCGCTAT CTGGATCAGT ATGAGGATGT	1500
GATTCTACGG GAAATTAAGG CTCAATTTCC AGATGTTGCA GTTGATAAAC TCATGGAAGA	1560
GTATATAAAG GCAGGCTTGA TTCTACGTGA AAATAAGCGC TATTACCTCA ATTTTCCTAC	1620
GCTTGAATCA CTTGATAGTC TTGAACTGGA TCAAGAGATT TTTGTCAGAG AAGCTAGTCC	1680
GGTCTATCAA GCCTTGTTGG AGCAGAGTTT TGAGACGGAA TTGCGCAATC AAATCAATGC	1740
AGCTATTTTA GTTGAAAAGA CGGACTTTGC GCGCATTAAT ATGACCCTGT CCAATTATTT	1800
TTACAAGGTC AAACAGCAGT ATCCTTTGAC AGAAAAACAG CAGGAGCTCT ATGACATTTT	1860
AGGAGATGTT AATCCTGAGT ATGCCCTCAA GTATATGACG GCTTTTTTGT TGAAATTTCT	1920
CAAAAAAGAC CAGCTTATGC AGAAATGCCG TGATATCTTT GTGGACAGTT AGGTTGTCTT	1980
AGGCTATATT GTGCAAAATG AAGATGGAAA GTATGAGTTG GCTATCGATT TTGATAAGGA	2040
GAGGTAACT TTCTACTTAG CGTGATTTCT TGTTCCTGAG TACATTGTTT GACTTTCCTT	2100

1227

AGTATTCCGT ATAACTATA TGTAACCGGT AACACATATC GGAATAAACT AAAGGAGACA	2160
ATCATATGTC ACTTGAAAAC AAATTGGAAC AAGCAACAGG CGCTGTCAAA GAAGGTTTTG	2220
GTAAAGTTAC TGGAGACAGC AAGACAGAAC TTGAAGGAGC TGTGAAAAA ACAGTTGCTA	2280
AGGCAAAAGA CGTTGTAGAA GACGCAAAAG GTGCTGTAGA AGGTGCCGTT GAAGGTTTGA	2340
AAAACGTTTT TACTAAAGAA TAGGAAAAA TCAAGGGTTT CATTTTCCCT TGATTTTTTC	2400
TATTCTTATA AATAATTTTC TCGACGGCT GTATCTCTG GGTAGGATTC TTTCTTGCCC	2460
TGGATGATT GGTAAACAATC GGCTCCCTTA CCCGCAATAA TAACTGCATC TAATTCGTGA	2520
TTTGTGATAG CCATTGCCGC CTGATGGCT TCTTGGCGAT CCGCAATCTT TTCAACAGGA	2580
TGATTGATGT AGCTACTAAT TTCATCTGCA ATGGCCATTG GGTCTTCATA GTTAGGGTCA	2640
TCAGCAGTCA GAAAGACTTG AATCTCAGGG TGTGTATTGA GGAGGAGGCC AAAGTCCTTA	2700
CGACGACTTT CTCCCTTGTT TCCTGTTGAT CCCAGAACCA GAGCAATCTT TCCGGTTTGA	2760
TGAGTTTCAA CCACATTGAT GAGTTTTTTC AGACTATCCC CATGTGGGC ATAGTCGATG	2820
AAGACCTTGG CTCCATTTTT CTGAGTGAGG ACTTCCATAC GACCAGGAAC GCGGGTTGCA	2880
GCGATGCCTT TTTGATGTC CTCAAGACTT GCTCCGAGAC GGAGACAAGC AAGTCCAGCA	2940
GCAACTGCAT TTTCTTGTT GAAGTTGCCA ATGAGTTGAA TATCATAATC TCCAGCGAGT	3000
TTACCCGTAG CTGAAAAGCT AAAGGCTTTG GAATTCTCGA TTTGGTTATC AAATTGGCTA	3060
CCATAGAAAT CATGGTCTTG ATCTTCAACC TGTCTTTCA AGACTGAGAA GTGGTCCATG	3120
TCACTGTTAA TGATGACTGC TCGGCTCTTT TCCATCAAGA GACGCTTG TGAGAAATAG	3180
TCTTCAAAGC TAGGGTGTTT AATCGGGCCG ATATGGTCTG GGCTGATATT TAGGAAAAC	3240
CCCACATCAA AGGTTAGACC ATAGACACGT TTGACCAGAT AGGCTTGACT GGAGACTTCC	3300
ATGATGAGGT GGGTACGGTC ATTTTGCAACA GCCTGATTCA TCATGTCAAA GAGGTCAATA	3360
CTCTCAGGGG TTGTCAACGC TGACTTAAAG AAAGTCTCGC CATCAAGAGT TGTGTTTCATG	3420
GTCGACAACA TAGCAGGTCT ATGCCCTTGA GATAAGATGT TATAGGCGAA ATAGGCTGCT	3480
GTTGTCTTAC CCTTAGTACC AGTAAAGGCA AGGAGTTTGA GTTTTTCTG TGGATTACCA	3540
TAGAACTCCA TGGCAATCAA ACTCATGGCT TTCTTTATAT CGTTCACAAT GATGACAGGG	3600
ATACCGACTT CGTAGTCTT TTCAGCTACA TACCAAGCTA ATCCTTGTGT TATAGCAGAA	3660
AGAAGGTATT CTTTTTAAA GGCAGCCCT TTTGCGAAAA AAAGAGTGTC TTCTGTACT	3720
TTTCGGCTGT CGTAGCTGAT GCTATCAAAA ATAACTTTGC TGTAGTTGTA GTGGTAATGA	3780
CCTTGGTCAA TAATTCGCG AAAAAGGCCA TCTTTCTTTA AAATATCTAA TACGGTTTCA	3840

1228

ATCTTAATCA TACTTTCTAT TGTAACCGA AAGTCGTAAA TTTACAAGTA ACAAGGAAAA	3900
GTTTATAATG GAAGATAAGG AGTTTTTCCT AGTTATCAAA ATTGAATGAG GAATCTATGT	3960
CGCACGAAAA CAATCACCAG CAGGCCCGAGA TGTTACGGGG GACTGCTTGG CTAACGGCTA	4020
GTAACCTTAT CAGTCGCCTA CTCGGGGCTG TTTACATTAT CCCTTGGTAC ATCTGGATGG	4080
GGGCTTATGC AGCTAAGGCA AATGGTCTCT TTACCATGGG TTACAATATC TATGCTTGGT	4140
TCTTGTTGGT TTCAACAGCG GGGATTCCAG TTGCGGTGGC CAAGCAAGTT GCCAAGTATA	4200
ATACCATGCG AGAAGAAGAG CATAGCTTTG CCCTGATTCTG GAGCTTCTTA GGCTTTATGA	4260
CAGGACTAGG CCTGGTTTTT GCTTTAGTCT TGTATGTCTT TGCTCCTTGG CTAGCAGACT	4320
TGTCTGGCGT GGGCAAAGAC TTGATCCCAA TCATGCAAAG CTTGGCTTGG GGAGTCTTGA	4380
TTTTCCCGTC TATGAGTGTT ATCCGAGGAT TTTTCCAAGG GATGAATAAC CTCAAACCTT	4440
ATGCCATGAG CCAAATTGCT GAGCAGGTCA TTCGTGTTAT CTGGATGCTC CTAGCAACCT	4500
TTATCATTAT GAAGCTCGGT TCAGGAGATT ATCTAGCAGC CGTTACCCAA TCAACCTTTG	4560
CTGCCTTTGT CGGTATGGTA GCCAGTTTTG CAGTCTTGAT TTATTTCTTT GCCCAAGAAG	4620
GTTCACTCAA AAGAATCTTT GAAACAGGAG ATAAGATTAA CAGTAAGCGT CTCTTGGTTG	4680
ATACCATTAA GGAAGCCATT CCTTTTATCC TGACAGGGTC TGCCATCCAG CTCTTCCAGA	4740
TTTTG	4745

## (2) INFORMATION FOR SEQ ID NO: 219:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1900 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

CCTGATTGAC CTTATAATAA GGAACAAAAC ACAATGCACT ACCTTTTCAA CAAAAGAGTT	60
GCTGCTTGAT TAAACCATC ACACCACTTA TACCATTTTG CTTTATACCC ATCTTGAGCT	120
AGGATACGAT CTTCTAAATC AAAACAGAG TAAATCTTTC TTTCTCGCA AGCTTGCGCA	180
TAGAGATGAT ATAGTTCATC ACCACCATCT CTATCCCACT CAGCAGAAAT CGTATCCCGA	240
CCTGCCAATA AAGCCTGATA AGCCCTGTGA TGCCCATCTG TAATCAGCAA ACAATCTCCA	300
AAGGCAAGAA TACTGATTGG ATCGACTTGG ATTGTTTCTG CCGACTGGTA AAGCATCTGA	360
ATATCTTGCA ACTTCTTTTC TGATAAATAT AGTTGAGTCA GATGAAGATC TGCTATATTG	420
ACTTTCATTT CTTTCTCTTC AAGGGAATTC GATACTCACT TCTGTTTGCC TTAAATCGC	480

1229

CATTGGAAGC GGAgCTTGTC ATAAAAGGGA AACTCGATAA ACAGGACTCC CAAGCCCACA	540
CAGAGACTGG CAAGGACGTC TGATGGGTAA TGAACGCCA GATAGACTCT TGATACCAGC	600
ACACTGACTA GGTAGAGGCC AAGGACGATT TGTACGATTT TTCTCCAGAC CTGATCTTTA	660
ATCCGCTGAC TAAGAATAAC AATCAAAGTC CCTACCATCA GCGTTACAGC TAGAGAATGC	720
CCACTTGGGA AGGAAAATCC CTTCTCCTCC ACCAGATGTA AAATAGCTGG TCGTGGGCGC	780
TGGTAGATAT TTTTAAAGGT CACGATTAAA AGACCTGCCA AAGCCAGATT TCCCAGCATG	840
AAGAACTTT CTATCTTCCA TCGCTTACGA TAAAAGACAA AAGCTGTAAT GACAACCCAA	900
GTGATAATCA CTGGGATATC AATCAGACGT GTGAGGGCTC GAAAAAGAA AGTCAAATAA	960
TCTGGTAAGT CTCCTCGAAT GGCAGTCTGA ATCGATTGGT CAAAATTGAC CAACATTTCA	1020
GGGTAAAATT TGACCATGTA GCCAAGAATA ACGAAAAGTA AAAGGGCAAA ACTGCCCTTC	1080
ATTAAAAATG TTTGTTTATC TCTCATAATG TTTAAGGTT GGTTTCAAGA GAACATACAA	1140
CAACCAGAAT GAAACGGAAA AGATAACACC TTCAATCAAG TTAAGGTA ATACCATGGT	1200
CATTAGGTAG TTGGAAAGTC CCAAAATTTT TCCAATATCA AAGTTAGCAA ACTTAGCGTA	1260
CAAAGGAACA GCATAAACAT AGTTGAGAAC CAACATGGCC AAGGTTAAAC CAATAGTTCC	1320
AGCTAGAGAG CCTAGTAGGA AACGAAGGGT TGTCGGTTCC TTTTCCAAA TCAAAGCAAA	1380
TACGATGACA AAAACTCCCA AAGCTACGAT ATTTCATCGGC AAACCAATGT AAGTATTCAC	1440
TCCTTGGCTG TTAAGAAGCA ATTTCAAGAG TGAGCGAAGC AAGAGCACTC CTAGAGmCsC	1500
AGGCAATCC ATGACCACCA GACCCACAAG GACTGGCAAG ATACTAAATT CGATCTTGAG	1560
GAAAGATGCC GCTGGTAAAA GCGGAAAGTC AAAGTACATC AGCACAAATG AGATGGCTGA	1620
TAGAATTGCA ATGGTCGAAA GTCGACGTGT GTTTGTCATA ACAGGTTTCCT CCAATTTTCT	1680
ATAAATCAG AAGAAGTTGG AAAGGATTCC TCTATCTATT CTCACCTTTT ATATCCCCAA	1740
AGTTCCCTCT TACTCTATTA AAGAAAAACA AAGCAAGTGG TTACAATCCG GCTATAAATC	1800
TATCAAAACA GACAAGGCTA TTCTTTCGTC TTCTCCCATC CAGACTATAC TGTGCGTTGT	1860
GGAATCTCAC CACATCACGT TGCCTCACG GACTTCTTTA	1900

(2) INFORMATION FOR SEQ ID NO: 220:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4692 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

GGTTTTCCAG CAGGAGCTTC TCCTTTATCA GAATGACCAT CCCATCTGCT CACGATAGAT	60
GAATAATGAT ATTTTTTACC ATGATAGTAA TTTGAAAAAG CCTAACCACC TCCTGAACCT	120
TCTCCATATG TCCATACTCC TCCATCTGGA TATTATACAG CAGCTGATGC AGCTCCCAAT	180
AATGTAAAC TTGAAATAAG AGCTAGAGCA AGTAATCTAT GTTTTTTCGT TTTTCATTTA	240
TTTTTTCTTT CAAAAAAGC ACACCTTGAG CAACAATGCA ACAAATAAA TCCTCCTCTC	300
TCTTTTATTG AAACCGCTTT CTTATGTGAT AAGAATAACT TTTTATCAT TTGTTGTCAA	360
GGAAAAATC GAATTTTTTA GATATTTTAC TATATTACCT CTGTGAATAA TATTATATAG	420
TAGTTTTATT TCAAAATAAT ATGCAACCAG TACTAACCAA ATATAAAATA GATGCCATTA	480
ACGAATTTTA TTCAAGTTT TCCCATTCAT ACTATACAAG TAAAAGAGAT GGTGTTAACT	540
AAAAAGCAAT TCAAACTATT GTAAAATTCC TAGCAAAAAG AGAGCCGAAA CTCTCTTTTT	600
TATCTTCTTT TACTTTTTTT GACTGGCATG AGTGTGATGT CTCTAACACT AAAGTAAGCT	660
AGGATCAACA TGGCTATTGC TAGGAATATT TCTGTTGGTA ATTGAAAAAT TTTCAGAAAA	720
GATAGAACCA ATAAATCAA GAGTGCCACT AAAATACATA CCATAGCGAC GATATTGACA	780
GTCCCTTTAA TGCTTCTGG TGTGCAAAAT ACATAGAGTA GGAGCAGTAA AATTCCTAGG	840
ACTAAATAGA CCATCTTCT CTCTTCTAG CTCTTATTCA GCTGATTTTT TCTTCTGT	900
AGCTTTCTCA CGCTCTGCTT TGTTAAGGAT TTGTTTACGC AAACGGATAG ACTCAGGCGT	960
TACTTCCATG TACTCATCGT CGTTCAAGAA CTCAAGAGAC TCTTCAAGTG TCAAGATACG	1020
AGGCGTCTTG ATAACAGCTG TTTGGTCCTT AGTAGCTGAA CGAACGTTGG TCATTTGTTT	1080
TGCCTTCGTG ATGTTAACTG TCAAGTCATT TTCACGAGAG TTTTCACCGA TGATCATTC	1140
TTCATAAACC TCAGTACCTG GGTGACAAA GATCGTACCA CGTTCTTCGA TAGACATGAT	1200
TGAGTAAGTT GTAGCCTTAC CAGCATCGAT AGAAACAAGG GCACCACGGT GACGTCCACC	1260
AATTTCCCTT GGAATCAATG GCAAGTATTG GTCGAAGGTA TGGTTCATGA TACCGTAACC	1320
ACGAGTCATT GATAAGAACT CAGTTGAGTA TCCAATCAAA CCACGCGCTG GAACAAGGAA	1380
GACCAAACGA GTTTGACCAT TACCAGTTGA AATCATATCC AACATTTTAC CTTTACGTTT	1440
AGAAAGGCTT TGGATAACAG ACCCTTGGTA TTCTTCTGGA GTGTCGATTT GTACACGTTT	1500
AAATGGTTCA CATTTAATAC CGTCGATTTT TTTTACGATA ACTTCTGGAC GAGATACTTG	1560
AAGTTCATAG CCTTCACGAC GCATTGTTTC GATAAGGATT GACAAGTGCA ATTCTCCACG	1620
TCCTGAAACA GTCCATTTAT CTGGTGAATC AGTTGGGTCA ACACGAAGGG AAACGTCTGT	1680
TTGCAATTCT GCCTGCAAGC GTTCTTCCAC CTTACGAGAA GTTACCCATT TACCTTCTTT	1740

1231

ACCAGCAAAT GGTGAGTTGT TGACCAAGAA AGTCATTTGA AGAGTTGGCT CATCGATGTG	1800
TAGGATTGGA AGAGCTTCTA CTGCATCTGT CGGAGTGATG GTTTCACCGA CAAAGATGTC	1860
TTCCATACCT GAAACGGCAA TCAAGTCACC CGCTTTGGCT TCTTGGATTT CACGACGTTT	1920
CAAACCAAAG AAACCGAAGA GTTTTGTAAAC ACGGAAGTTT TTAGTTGTAC CGTCAAGTTT	1980
AGAAAGGGTA ACTTGGTCCC CAACCTTAAC TGTACCACGG AAGACACGAC CGATACCGAT	2040
ACGTCCAACG AAGTCATTGT AGTCCAAAAG TGACACTTGG AACTGCAAAG GCTCATCTGA	2100
GTTATCTACT GGAGCTGGGA TATGGTCGAT AATCGTGTCA AAGATTGGTG CCATAGTCGC	2160
TTCTTGGTCA GCTGGATCAT CTGACAATGA AGAAGTTCCG TTGATCGCTG AAGCATAAAC	2220
CACTGGGAAA TCAAGCTGGT CGTCATCTGC ACCAAGCTCG ATGAAAAGTT CCAAGACTTC	2280
ATCCACTACT TCTGCTGGAC GAGCTGATGG CTTATCGATT TTGTTAACAA CCACGATTGG	2340
GACAAGGTCT TGTCCAAGG CTTTTTCAA TACGAAACGA GTTTGTGGCA TGGTTCCTTC	2400
ATAGGCATCT ACGACCAAGA CAACACCGTC AACCATTTTC ATGATACGCT CAACTTCTCC	2460
ACCAAAGTCC GCGTGTCTG GTGTGTCCAT AATGTTGATA CGAGTTCCGT TGTAAGCAAC	2520
GGCAGTATTT TTAGCAAGGA TGGTAATTCC ACGCTCTTT TCGATATCGT TTGAGTCCAT	2580
AGCACGCTCT GCCAATTCAAG TCCGTGCATC AAGCGTTTCT GATTGTTTCA ATAATTCTGC	2640
AACCAGGGTT GTTTTACCGT GGTCAACGTG GCGGATAATC GCAAATGTTAC GGATATCTTC	2700
TCTTAATTTT GTCATGATTT CCTCTATAAT ATTCAAAATT TATTTTCTAA CTGAACGATT	2760
ATACCATAAT TTCAAATAAA TAACATAACT CAAGCAAGTG TAAATGTTTT CACTCTGCTT	2820
TTCTTTTCAC GTCAAGCCTT TTCAAAGCGA GCGACTTATG ATAAGATAGG CACAGTATGC	2880
GTTTAGATAA TTTATTAGCT CAAGAAAAAA TCAGCCGAAA GGCCATGAAG CAAGCACTCC	2940
TCAGAGGGGA AATTCTAGTC GATGGTTGCC CAGCCCCTC CCTAGCTCAA AATATCGATA	3000
CAGGACTACA AGAACTCCTT TTTCAGGATC GAATCATTCA AGGCTATGAA CACACCTATC	3060
TTATGCTTCA TAAACCTGCT GGTGCCGTTA CAGCCAACAA AGACAAGGAA CTTCCGACCG	3120
TCATGGACCT GCTTCCATCT AACATCCAGT CTGACAAGCT CTATGCCGTT GGCCGACTGG	3180
ACCGAGATAC AACGGGACTC CTCCTCTTGA CCGATAACGG TCCCTTGGGC TTTCAGCTCC	3240
TCCATCCCCA ATATCATGTC GATAAGACTT ACCAAGTTGA GGTAAATGGA CTTCTAACAC	3300
CTGACCATAT CCAAACCTTT CAAAAAGGAA TTGTCTTTTT AGATGACACT GTCTGTAAAC	3360
CCGCAAAACT AGAGATTCTA TCTGCAAGTC SCTCCCTCAG TCAAGCCTCT ATCACCATTT	3420
CAGAAGGAAA ATTCATCAA ATCAAGAAAA TGTTCTCTC GGTGGTGTT AAGGTGACTA	3480

1232

GCCTCAAAAG AATCCAATTT GGGGACTTCA CATTGAACCC AGATTTAGCA GAAGGTAAC	3540
ACCGCCCTTT GAACCAAAAA GAGTTACAAA TCATTAAAA CTATTAGAG ATGAGTCGAT	3600
AAAACAAAA AAGCTTTAAA ACTAAAGCTT TTTTCTTTA TTTACCGAAA AATTAAGGCG	3660
ATTGCTACAA TCCAGTTAAC TACAGAAATC ACAATTCCTA AGATATTAAG AATCTTTCT	3720
ATTTTATAGT CTAATTGTGA CTCTTTTGG TATGAAATAG CCAAGACCAA TCCTATGATA	3780
CCCAAAATCA GGCCTACAAT TGGAAATAAC AAACCAAGAA TAATCGACAA GATACCCACA	3840
AAAAGTGGAT TTTTCTTCTT TTCTTTTATG TTCTAAGAAC TCCTTAAATT TTATACAAAT	3900
TAATTATACT ATAAAAAAT AGCTTCATCC TATCATTCGA CTAATTTGGA AATAAGGTTA	3960
GCTAGTCTTC ACTTTCCTT TCCAAGAATC CAAGCCATAA GAAAGGATAT AAATCTCAGA	4020
AAAACCTTGT TTTTCAAGT AAAGAGCTGC ATTTGTAAC CGTTGCGCAC GTTGGTTTTC	4080
GTAGAGAAGG ACAGGTTTAT CTTTACGAAG GGCTGCAAGA CTAGTTTCA ACTGACTTGA	4140
AGGAATATTG CGTGCACCAA GGATATGTTT TCTGTGGAAT TCTGCTGGGT CGCGCAAATC	4200
AATCAATTGA CCCGTACGAA TCAAGGCTTC AAATCCTCA TTGTCCACAA TTTTAGCCGC	4260
ACGGCGAATA CGAAGATAGT TAAAGCCCAT CCACGCCAAC ATTGCTAGTA TAAGTGCCCA	4320
CAAAATCCAA GTAACCATTA GTTCTTTTCT CCATTTTCT CAATATAATC CAATTCTACC	4380
TTGTGCTCTC TCGGAAGAAC TGCTTCTGCC TCTAGATAGT CTAATTTATC CATCAACCCT	4440
GCATCGTAAA TCCGAGATAG TTCCAACCTC ATCAGTTCAA TATCATATAA GCGTTTTC	4500
ATGTAAACAA TAATACCAA TCGTTGAGG AATTGCTGCA CATCATAGAA TGTTTTCATA	4560
AGACTCATT TAGCAAAAT TTGTGTTTTT TTCAAGAAGA GACTCACACA ATGCTCCTTA	4620
TTTTCCTATC TTCTTAGCG ATTCTAAGGC AAGTATGGTA CAATAAAAC ATGGGGATTC	4680
AACAATTACA TT	4692

(2) INFORMATION FOR SEQ ID NO: 221:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 706 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

GCTAAAAAGC TGATAATCTT CGACTCCTGT ATATGATGTG TCTTTTCATG TAAGACACGC	60
GCCGCCAGAA TCATGGCAAG AGCTGCAAGA CTGGCAAGTA AGAAGCCGAT AAGATAGGCA	120
AAAAGATAAG TGAATTTGAC AAAGAAAGTC AAAAGAACTA GGAAACCAAA GCCTCCTCCA	180



1233

AAAACTACCA AAGTCTTTCG TAAATCCCAG ATTTTATCCA ACTGCTTGAC GAGGGAAGTC	240
GTCTGACGAA CGCCTACAAT AGTTGCTAAC ATACTTCCTA AAAAGAATGG ATAGACATGA	300
GTTAAACTGG AGAAATAAAC AGAGGAATAA GAGGTCACTA GAAACTACC AATAAACATG	360
GAGAAGAAAC TGATCAAGAA GGCAACAGCA GATAAGAGAA AGACCATCCC CTTCAACTGA	420
CCATTTGATT TAGCTTGTTT GGATAAGAAC CAACTGCCA ATCCCCAAAG AATATAGTAG	480
TGAACCTCAA CTGCCAACT CCAATTATGA ACAAACAAAT GAGGAATGAA CTGAGATTCA	540
TAACTCCAC CTGTTAGGAG TTCATAGAAG TTGGTCATA AGCCTAAGAC GCCCGCAATC	600
TGGCCACCAA TTCCAGCAAC ATAGTCTTGG CGAACCAAGA AAGTAAAGG CATGGTCACC	660
AAGACCATCA AAACCACAGG TGGCACAATC TCGATAAAAG CGTCTT	706

(2) INFORMATION FOR SEQ ID NO: 222:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3236 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

CAGCTGATGG GCAATATCAG TCATAGAAAT TTTTCAATT AACTTTTGAG CAATTTTTTG	60
GTTGATGATA CGAGGGATTT GGTGATTTTT CTTTACCAGG GGAGTCTCAG CAACCATCAT	120
TTTTGAACAG TGATAGCACT TGAAACGGCG TTTTCTAAGG AGAATTCTAG AAGGCATACC	180
AGTTGTTTCG AGGTAAGGGA TCTTAGACGG TTTTGAAG TCATATTTCT TCATTAGACT	240
TCCACAATCA GGGCAAGATG GAGCCTCATA ATCCAGCTTA GCGATAATTT CTTTGTGGGT	300
ATCCATATTG ATGATATCTA GAATCTTGAT GTTTGGGTCT TTAATATCGA GCAGTTTTGT	360
GATAAAATGT AATTGTTCCA TATGATTCTT TCTAATGAGT TGTMTGTGCG CTTTTCATTA	420
TAGGTCATAT GGGACTTTTT TTCTACACAA AAATAAGCTC CATAATATCC ATAGGGGATT	480
TACCCACTAC AAATATTATA GAGCCCGAAA ATATGGGAAA ACTGATCCTT GTTCTGCTT	540
TTGTCTATAG AAGAATAATA AAGATTATCT TCTTCAAAT CTCCGATATT CTCTAAAGTT	600
TTGTGCAAGT TGCACAGAAC TTGTTTATTT TTTTGGTCAT CTTGCCATAG AAATATAAAG	660
CGTTTTTATA TATAATATAA TTATCAAAG ACAAAGGAG TTCACCTCAT GGTAGAATTG	720
AATCTTAAAA ATATTTACAA AAAATATCCA AACAGCGAAC ACTATTCAGT TGAAGATTTC	780
AACTTGAACA TCAAAGATAA AGAATTTATC GTTTTCGTAG GACCTTCAGG ATGTGGTAAA	840

1234

TCAACTACAC	TCCGTATGAT	TGCTGGTCTT	GAAGACATTA	CAGAAGGTAC	TGCATCTATC	900
GATGGCGTAG	TTGTCAACGA	CGTAGCTCCA	AAAGACCGTG	ATATCGCCAT	GGTATTCCAA	960
AACTACGCTC	TTTACCCACA	CATGACTGTT	TATGACAACA	TGGCTTTCGG	TTTGAAATG	1020
CGTAAATACA	GCAAAGAAGA	CATTAACAAA	CGTGTTC AAG	AAGCAGCTGA	AATACTTGGA	1080
TTGAAAGAAT	TCTTGGAACG	TAAACCAGCT	GACCTTTCAG	GTGGTCAACG	TCAACGTGTT	1140
GCCATGGGGC	GTGCGATTGT	CCGTGATGCG	AAAGTATTCT	TGATGGACGA	ACCTTTGTCA	1200
AACTTGATG	CCAAACTTCG	TGTATCAATG	CGTGCTGAAA	TCGCTAAAAT	TCACCGTCGT	1260
ATCGGAGCTA	CAACTATCTA	TGTAACCTCAC	GACCAAACAG	AAGCGATGAC	ACTTGCAGAC	1320
CGTATCGTTA	TTATGTCAGC	TACTAAGAAC	CCTGCTGGTA	CAGGTACTAT	CGGACGTGTA	1380
GAACAAATCG	GTACTCCTCA	AGAAGTTTAC	AAAAATCCAG	TTAACAAATT	CGTTGCAGGA	1440
TTTCATCGGAA	GCCCAGCTAT	GAACCTCATC	ACCGTGAAAT	TGGTTGGTAG	CGAAATTGTT	1500
TCTGACGGTT	TCCGTTTGAA	AGTGCCAGAA	GGAGCATTGA	AAGTTCTTCG	TGAAAAAGGC	1560
TACGAAGGAA	AAGAATTGAT	CTTTGGTATC	CGTCCAGAAG	ACGTGAATGC	AGAACCTGCT	1620
TTCTCTGAAA	CATTCCCAGA	CTGTGTTGTA	AAAGCGACTA	TCTCTGTATC	AGAACTGCTT	1680
GGTTCAGAAT	CTCACCTTTA	CTGTCAAGTT	GGTAAAGACG	AGTTTGTGTC	AAAAGTTGAT	1740
GCTCGTGAAT	ACTTGCAAAC	AGGTGCAACA	GTTGAGCTTG	GATTTGACTT	GAACAAAGCA	1800
CACCTCTTCG	ATGTAGAAAC	TGAAAAAACA	ATCTACTAAA	ATAAATAAAA	TTCAAAGCAC	1860
TACAAGAAAA	GATATCTCTT	TATCAATTGT	AGTGGAGAGA	TATCAGTTAA	TCTAGGGAGA	1920
GAAACAAAAT	GCTTCTCTCC	TTTTTGCTAG	AGAAGTCATA	TTATGCATCT	ATATTGTGAT	1980
GCTCTTTAAT	ACTCTTCGAA	AATCTCTTCA	AACCACGTCA	ACGTGCGCTT	GCCGTACGTA	2040
TGATTACTGA	TTTCGTCAGT	TTTATCTGCA	ACCTCAAAGA	TGTACTTTGA	GCAGCTTACG	2100
GCTAGTTTCC	TAGTTTGCTC	TTTGATTTCC	ATTGAGTATT	ATTTGTGGGT	ACCATCTACA	2160
AGTGAAGCTA	TATGCGTAAA	CTACGTGAGC	AATTGAATTC	GAAGTAGAGA	GGTAATAATA	2220
AATTTATGCT	ATAGTTATGG	TGACTTGTAT	GCTTTTGATT	CTAGTTTATC	AAATAATAGA	2280
TTAGAATTGT	CAGATAATAT	CATTTTGTGT	TATAATGAAG	AAAAAACAGA	GGTGTTCAAA	2340
TGTCAGAAGC	AGGTCATAAG	TTTTTAGCAA	AATTGGGGAA	AAAACGCTTA	CGTCCAGGTG	2400
GAAAGCGTGC	CACAGATTGG	TTAATTGCAG	AAGGAGGATT	TTCAAAAGAA	AAGAGAATAC	2460
TAGAGGTTGC	GTGTAATAGG	GGAACCTACAG	CAATTGAGTT	GGCACAGCGT	TTTGGTTGCA	2520
AGATAACTGC	TGTTGATATG	GATGCTCAAG	CTTTAGAAAGT	GGCTAAAAAA	TCTGCTGGAA	2580
CGGCAGGTGT	TGCTCATTTA	ATCAGTTTGT	AAAGAGCAAA	TGCAATGAAA	CTTCCTTATC	2640

1235

AAGATGCTAG TTTTGATATT GTTATAAATG AAGCTATGCT GACTATGCAA GCCGATCAAG	2700
CTAAGAAAAA ATGTGTAATG GAATATCTAA GGGTATTAAA ACCTGGAGGT CTTCTCTTGA	2760
CACATGATGT GCTTCTTAAG GAAGCTAAAG AGTCTATCAG ACAGGAATTA TCACAAGCAA	2820
TTCATGTAAG TGTAGGTCCT TTAAGTCAAG ATGGTTGGGA ACAGGTGATG ATAGAATCAG	2880
GTTATTGTGA TGTGAAAGCA TTGACTGGTG AAATGACATT AATGAAATTA TCGGGTATGA	2940
TTTATGACGA AGGTTTGCTA GGAACCTTGA AAATTTGTGT AAATGCTTGT AAAAAGGAGA	3000
ATAGAAAGCA GTTTTAACT ATGTATAAAA TGTGCTGCTA GAATAAACAG AAATGGGGCT	3060
TTATTGCGAT GGCTAGTTAT AAATCGTCAA AACGTTAGAT AATTATTGAA GTTAACTTTT	3120
CCTTTTTTCT TTCTTAAAAA ATATGCTATA ATAGAGAGTA AAAAAGTTTG AAAGAAAGAA	3180
AAAGATGAAT TTAAGAGATT ACATTGCAAC AATTGAAAAT TATCCAAAGG GTACCG	3236

(2) INFORMATION FOR SEQ ID NO: 223:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2885 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

CCTGACTTTT CAAATTGGTT AGTTTGCCAC ACTTGCTTGA TATGTCGTG GAAAGCATGG	60
CTATTACTTC TCAAAGGGCG ATTTCTCACC CCATGAAAAG TGTCTATTTT TGTTAGGTT	120
TGTAAGTTAA TTCATTGTCA CATATTACTC TTTAACTGAT TGAGTGAGTA CCGCTTATAT	180
TTGATGCCAA ACGCCTTAAA AGTGTACCC TCAAGTCCTT TTAGAATACG GCTATAATTC	240
CGCTCATTGT AAAGTATCTT AAGCTCATCA CTATCTAGGT TGGTATTAAA AATGGTATTT	300
TCACGATTGT TTAGCACGTC AAAGAGTAAA TCCTGCTCCC AGTCACTCTT AGGCTTAATA	360
ACAGCATTTT TTGCTCCTAA ATCATCAATA ATTAAGTAAT CAACAGACTT CATGAGTTCA	420
GTAGCTTCAA ACTCTGTAAG TGTGACACCT TTACCATAAT TCCACCCCTC TTAAATTGT	480
TTGATCATTT CGGTTAGGCT TACAAAAAGC AACTCTTAG GTTCTCCTTT TGTCTTATAC	540
CCCTCATTTA TACCTTTGGC AATAGCAACT GATAAAAGTG TTTTCCAAT CCCTGTACCT	600
CCTGTGATAA GCGTATTTCC CCTCATGCCA TCAAGATATT TTGTACCTG ACCTTTTGCA	660
AATTCTAAAA ATCGCTTTTC TTCTGATGTT ACAGCATTA AATCATCAAA AGTTTGTAGTT	720
TTAAACTCAT CTGCTACATA GCTCTTATG CTCATCAACA CATTATAAGT TTGCATATAT	780

1236

AGTTTAGCAT	TCAAATTATC	AGCAATCGCA	TCTTCTTCAT	CTTGCTTTTT	CTGTTCTTCT	840
TGGCATTGTT	CACAATAGGG	TGGGATACAG	CGAACTTCTT	TTATTGCCTC	TCCGTTCTCA	900
TTCCACCCCA	CTACTACATG	TCTTTCTCCT	TTGATTGTG	TTAGCTGTAT	TTCATGCTTA	960
GGACACAATT	CGTCTAGTTT	AAATGTCTCA	ATATTTCCTA	AACTAGATTG	TAATGATTTC	1020
ATTTTCTGAC	CTCCTAAAAT	GGTTTTTCTT	GTGTTGGTAT	CCAATCTTCA	TAGCTGGTAG	1080
GCTCTAGTTG	ATTGGTTTGC	TGTTTTTTAG	CCTCAGCGGC	TGCCCTGCTA	TTTCTAACAA	1140
GTTCCACCGT	CAATAAATTG	TCCTGTTTCC	AACGGTTAAG	GATTACCTTG	ATGTATGCAA	1200
AGTTTGCTTT	ACCCTGACTG	ACAGCCTCTT	TTAACGCCTC	ATGGATAAGC	TCTGGGTAA	1260
AATCTTCTAG	CATATACTGC	AATTCTTGAA	TCTGTAACGG	TGACAATGCT	TTACCTGTCT	1320
CAGCTCGCTT	CATATTCAAC	AAGTCGTCTA	TTTCCACACT	GGTTACTTTT	TTATTTACAA	1380
AATCAGAAAT	CAGTTGAAAA	ATGTTTGGAC	TTTGTAGCTG	GATTTTCAGCC	ATTACCTCAT	1440
CAAATCTGTC	TTGTGTCATG	TTGTCTAAAT	CTAGTGTCTA	TGCATTGCCT	CCTCAAACCT	1500
CTCTATAAGA	CAACTTTTAT	TTGCTTTCTG	AGTTCCATTT	TTAGAGTTAA	AAAGAATATC	1560
TTTAAAGGTT	ACAGTAGCCT	CTAAATACTC	CTTTTCAGCA	TGCTCTATAT	ACGCCTGTTG	1620
CTCTGCTTCG	TTCTCAAAAA	AGTGCTTAGC	TTGGCGTTTA	AAGAATGCTT	TTGCGCATAGC	1680
GTCCATTCA	AAAATACCAG	GGGCGAAAAA	CATTCCCGTA	GTGCTTTTAG	AGACCGCTTC	1740
GATTTTATGG	CTTTCATTCA	ATTTCAGGAAG	TTCAATCCAA	AGTAAACGGG	ACAACTCATC	1800
TTTGATGGAT	TTTGTCTGAC	TTTCCAATAA	AGAAAGGATT	CTTAGGCCAT	TTTCTTCGCT	1860
AATTTCTCGC	ATTTCTGCGC	TAATTCTGTC	TATACGTCTA	GTTAAATTCT	CATATGTTGT	1920
TTCTGTCTAG	TTTTTACCTC	TGTTTCTTTG	TTGGTGTGAT	TTTTTAGCTT	ATTTTTTTAC	1980
TTCTAAACAT	CATTGTCTTA	ATTTCTGAT	AACTCATTTT	CAATTCAATC	ATAGCTATTG	2040
CCATATCCTC	AAATGCCTGG	TACTGCTCCA	ACTCCTCACT	AGTCAAGCTA	TCGATACCGT	2100
TATAGCCCCC	ACGCTCTTCT	CTTAACTGCT	TAGCGTTCAT	GTCTGTTACT	GCCTTTAGTA	2160
GCAAGTTGTT	CATGGTGCTA	TGCGCGTGCT	TTGGTGCATT	AGGCCATGTT	TCTATACTGT	2220
CATGCAAGGT	TTTTCTTTTC	GGTTTTTCTA	GCGCCCTCTG	CAGACGAATT	TCAGAAAGTT	2280
CCTCAGCAT	TTCAAAGAAT	GCTTTGACTA	GGTTTAGTTT	GAATTGCCGT	ACTGTTTCGG	2340
TATTCCTTAA	ATAAGTGATC	AGAAAAGTAG	CCTGTTGCTC	GTTCAGAATA	TAGGATTTTT	2400
TAGGTTGTCC	TCTAGTATCT	AATTTATGGA	TTTTAAATCC	AAGTATPCCC	AACTCTTCAA	2460
AGTCAGCCTT	ATTTCTCTTT	ATTAAGCGCG	TGATAGTGTG	GTGTTGTACT	TCAGCACATT	2520
CAGCGATGAT	CTCGCTTGTTG	GTGTACGGCT	CTTTCTTACC	GTCCATGTAA	ACTAGTTCCA	2580

1237

TTACGGTTCT ACCTCCTGTA TAAATCTGGT TAGCTTACTT TTTAATTGCC TCCTCTAGCC	2640
TCTTTTTTAG CCTCTAAAAC GGCTTTGGCT AGTGGTTAAT ATTATTTACC ACTTGTCTCT	2700
ATAACGTGT TAGAGGCCTT TATAACGACT TGTATCGCTG TATCGATATC CTCCGTGGAA	2760
TAGTAGATTT ATTTTCTAAT ATCATTCAAG ACTTGTTTAA CCCATTCTT GAAAGAAATA	2820
AAATTACATC TTCTTTATCC TTGGCATCTG CTTTGTCTGA GACAAATTAG AATGTCAATA	2880
CTTGG	2885

## (2) INFORMATION FOR SEQ ID NO: 224:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3144 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

TATCAATCCT TTCCCATTAT AGGAGCAACA GAGTGGGAGT AGTCATCTAA GGACTAATTT	60
ATGTATTTTT ACGAGTCAGT ATCTTGGGAT ACTGGTTTTT ACTTTTCTAG ACTTTTGTAC	120
TACTTGTTAA AACTGGGATA ATTTTCGACT GTTTAACAGT TATTATGCAA AGTCTAAAAG	180
ATTAGAATTG TCAAAACAAT CCGTCTAGGC TTGATTTTAT CCTTTATTTA CTATAAAATG	240
AGAAGGAAAA ATGTCAAAC TTTATATTGC AAATAGGAGA AATCATGACA AAAACATTAA	300
AACGTCTGA GGTTTTATCA CCTGCAGGGA CTTTAGAGAA GCTAAAGGTA GCTGTTTCACT	360
ATGGAGCAGA TGCTGTCTTT ATCGGTGGTC AGGCCTATGG TCTTCGTAGC CGTGCGGGAA	420
ACTTTACTTT CGAACAGATG GAAGAAGGCG TGCAGTTTGC GGCCAAGTAT GGTGCCAAGG	480
TCTATGTAGC GGCTAATATG GTTATGCACG AAGGAAATGA AGCTGGTGCT GGTGAGTGGT	540
TCCGTAAACT GCGTGATATC GGGATTGCAG CAGTTATCGT ATCTGACCCA GCCTTGATTA	600
TGATTGCAGT GACTGAAGCA CCAGGCCTTG AAATCCACCT TTCTACCCAA GCCAGTGCCA	660
CTAACTATGA AACCCTTGAG TTCTGGAAG AGCTAGGCTT GACTCGTGTC GTTTTAGCGC	720
GTGAGGTTTC AATGGAAGAA TTAGCTGAGA TCCGCAAACG TACAGATGTT GAAATTGAAG	780
CCTTTGTCCA TGGAGCTATG TGTATTCAT ACTCTGGACG TTGTACTCTT TCAAACCACA	840
TGAGTATGCG TGATGCCAAC CGTGGTGGAT GTTCTCAGTC ATGCCGTTGG AAATACGACC	900
TTTACGATAT GCCATTTGGG AAAGAACGTA AGAGTTTGCA GGGTGAGATT CCAGAAGAAT	960
TTTCAATGTC AGCCGTTGAY ATGTCTATGA TTGACCACAT TCCAGATATG ATTGAAAATG	1020

1238

GTGTGGACAG TCTAAAAATC GAAGGACGTA TGAGTCTAT TCACTAyGTA TCAACAGTAA	1080
CCAACTGCTA CAAGGCGGCT GTGGATGCCT ATCTTGAAAG TCCTGAAAAG TTTGAAGCTA	1140
TCAAACAAGA CTTGGTGGAC GAGATGTGGA AGGTGCCCA ACGTGAAGTG GCTACAGGAT	1200
TTTACTATGG TACACCATCT GAAAATGAGC AGTTGTTTGG TGCTCGTCGT AAAATCCCTG	1260
AGTACAAGTT TGTCGCTGAA GTGGTTTCTT ATGATGATGC GGCACAAACA GCAACTATTC	1320
GTCAACGAAA CGTCATTAAC GAAGGGGACC AAGTTGAGTT TTATGGTCCA GGTTCCTGTC	1380
ATTTTGAAAC CTATATTGAA GATTTGCATG ATGCTAAAGG CAATAAAATC GACCGCGCTC	1440
CAAAATCCAAT GGAAGTATTG ACTATTAAAG TCCCACAACC TGTTCATCA GGAGACATGG	1500
TTGAGCTCT TAAAGAGGGG CTTATCAATC TTTATAAGGA AGATGGAACC AGCGTCACAG	1560
TTCTGTCTTA ATGTAGTTGT TTAGTTTAA AAACTATGC AAAGCTCCAT ATACAACACT	1620
TAAACGAGAT TAAAGAATGG CGAAATCCCT TGATGCGCAA GAGATTAGCT GTCTTTTITA	1680
TTTTTTAAGT GATAAAGTCG GAGTTTAGGC ATCAAAGCCT ATCAAATTAA ACAAAGAAGC	1740
GATGTCTTAG ATATTTTGAA AAAAATTAAT AAGCAGAAAA CTCTCTATTA TTTTGTGTGA	1800
GAGAGTTTTT TGTTAATAAA ATTTCAAAA ATGACATTTA TATATTGCAT TAAGTTAGAT	1860
ATATGATATA ATATTGTTAA AAAGAGGCGC AACTTTTAA AATTAATGAG AATCAAAGAG	1920
AAAACCAATA ATATTAATGG AGGAATAAAA AATGTAAGTA AGCATTATGG TCATTCAATC	1980
ATTCTCAAAG ATATAAATTT TGCACCTAAC AAGGGTGAAA TTGTTGCTCT AGCAGGGAGA	2040
AATGGAGTTG GTAAGAGTAC GTTGATGAAA ATTCCTGTTC AGAATAATCA ACCGACTTCA	2100
GGTAATATTA TAAGCAGTGA TAATGTTGGG TATTTAATCG AAGAACCAAA ATTATTTITA	2160
TCTAAAACAG GTTTAGAGAA TTTAAAATAT TTGTCAAATT TATATGGTGT TGACTACAAT	2220
CAAGAAAGAT TTAGATGTTT GATCCAAGAG TTAGATTGA CTCAGTCTAT TAATAAAAAA	2280
GTAAAGACCT ATTCTTTGGG TACAAAACAA AAATTAGCTT TGCTTCTAAC TCTCGTTACG	2340
GAACCTGATA TATTGATTTT AGATGAACCG ACTAATGGTT TAGATATTQA ATCATCACAA	2400
ATAGTTTTAG CGGTTCTAAA AAAATTAGCT TTACATGAAA ATGTGGGAAT TTAAATATCG	2460
AGTCATAAAT TAGAAGACAT TGAAGAAATT TGTGAGAGAG TTCTTTTCTT GGAGAACGGG	2520
CTTTTGACAT TTCAAAAAGT AGGAAAAGAT AGTCATAATT TCTTGTTTGA GATAGCTTTT	2580
TCATCAGCTA CAGATAGAGA CATTTTCATT ACCAAACAAG AATTTTGGGA TATTGTTTAG	2640
GAAGAGGGAT TGAGAATTAC TATGTCTGGG AATATTCAAA ATAGTGAGCT TTTTAAATTT	2700
TTTAACGAAA ACTCTATTAA AGTAGTTGAT TTTGAACTA AAAAAGAGAC GCTTAAAGAT	2760
ATTTACCTAA ATCGTTCAAA ATAAAGGAAG GTTATAATCA TGAAATTAAA TAAACAGAAG	2820

1239

AATCGGATGA TTTACGTCTT GTCTAATTTT CTATATGCTA TCTCAGTTTC CATTATTTAT	2880
GCTTTGAATG GCATTGTGTT ACTAGTCATA GTAAGTAAAT TGGGTATTCC AGGTGATTTA	2940
GGATTAAATT TTATAGTAGC TATTGTAGTC AATACAATTT TGTTAGTCCT GTTTTATTTT	3000
CTATTATCTT ACATTTTCTA TTTATACAAA TTGAAAAGTG GCTTGGTATW TGGTATTTTA	3060
GTAGCTTTAC TACTCTTTAT CTCTAATATA TTAAATACGA TGATGATGAA TACTAGTAAT	3120
GATTTGTTTA TCAAAGCAAT TGAA - - - - -	3144

(2) INFORMATION FOR SEQ ID NO: 225:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3766 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

TACGGTATTA TTTTAAAGGA GAAAGAATCA TGAAAATCAA AAAATGGCTT GGTCTAGCAG	60
CCCTTGCTAC AGTCGCAGGT TTGGCTCTTG CAGCTTGCGG AAACCTCAGAA AAGAAAGCAG	120
ACAATGCAAC AACTATCAAA ATCGCAACTG TTAACCGTAG CGGTTCTGAA GAAAAACGTT	180
GGGACAAAAT CCAAGAATTG GTTAAAAAAG ACCGAATTAC CTTGGAATTT ACAGAGTTCA	240
CAGACTACTC ACAACCAAAC AAAGCAACTG CTGATGGCGA AGTAGATTTG AACGCTTTCC	300
AACACTATAA CTTCTTGAAC AACTGGAACA AAGAAAACGG AAAAGACCTT GTAGCGATTG	360
CAGATACTTA CATCTCTCCA ATCCGCCTTT ACTCAGGTTT GAATGGAAGT GCCAACAAGT	420
ACACTAAAGT AGAAGACATC CCAGCAAACG GAGAAATCGC TGTACCGAAT GACGCTACAA	480
ACGAAAGCCG TGCCTTTTAT TTGCTTCAAT CAGCTGGCTT GATTAAATTG GATGTTTCTG	540
GAACTGCTCT TGCAACAGTT GCCAACATCA AAGAAAATCC AAAGAACTTG AAAATCACTG	600
AATTGGACGC TAGCCAAACA GCTCGTTCAT TGTATCAGT TGACGCTGCC GTTGTAACA	660
ATACCTTCGT TACAGAAGCA AAATTGGACT ACAAGAAATC ACTTTTCAAA GAACAAGCTG	720
ATGAAAATC AAAACAATGG TACAACATCA TTGTTGCAA AAAAGATTGG GAAACATCAC	780
CTAAGGCTGA TGCTATCAAG AAAGTAATCG CAGCTTACCA CACAGATGAC GTGAAAAAAG	840
TTATCGAAGA ATCATCAGAT GGTTTGGATC AACCAGTTTG GTAATAAGAA ACAGGGAGGT	900
GGGAGAGAAA ATTCCACCTC TTGCTTTTGT ATAGAGTATA GATTGTAAAG AAGACTATT	960
GTTCATAGAA AGGTAGAGAG AATATGGTTT TTCCTAGCGA ACAAGAACAG ATTGAAAAAT	1020

1240

TTGAAAAGGA TCATGTAGCC CAGCATTATT TTGAGGTTTT GCGTACCTTG ATTTCTAAGA	1080
AATCAGTCTT TGCCCAGCAG GTTGGACTCA AGGAAGTCGC AAATTATCTG GGTGAGATTT	1140
TCAAGCGTGT TGGAGCTGAA GTGGAGATTG ATGAGAGCTA TACAGCGCCC TTTGTCATGG	1200
CACATTTCAA GAGTTCGCGT CCAGATGCCA AGACCTTGAT TTTCTATAAC CACTATGACA	1260
CTGTGCCAGC GGATGGGGAT CAGGTCTGGA CAGAGGATCC KTTTACGCTT TCGGTCCGCA	1320
ATGGCTTCAT GTATGGGCGT GGGGTGATG ACGACAAGGG TCATATCACA GCTCGCTTGA	1380
GTGCTTTGAG AAAATATATG CAGCACCATG ATGATTTACC TGTCAATATC AGCTTTATCA	1440
TGGAGGGAGC GGAGGAATCG GCTTCAACAG ACCTAGATAA GTATTTGGAA AAGCATGCAG	1500
ACAAACTCCG TGGGGCGGAT TTGTTGCTCT GGGAAACAAG GACCAAAAAT GCCTTGGAAC	1560
AGCTGGAAAT TTCTGGTGGC AATAAGGGGA TTGTGACCTT TGATGCCAAG GTAAAAAGCG	1620
CTGATGTGGA TATCCACTCG AGTTATGGTG GTGTTGTGGA ATCAGCTCCT TGGTATCTCC	1680
TCCAAGCCTT ACAGTCTCTT CGTGCTGCGG ATGGCCGTAT CTTGGTTGAA GGCTTGTACG	1740
AAGAAGTACA AGAGCCCAAT GAACGAGAAA TGGCCTTGCT AGAAACTTAT GGTCAACGAA	1800
ACCCAGAGGA AGTTAGTCGG ATTTATGGAT TGGAGTTGCC TCTCTTACAG GAGGAGCGGA	1860
TGGCCTTTCT AAAACGTTTC TTTTTCGATC CAGCGCTTAA TATCGAAGGA ATCCAGTCTG	1920
GTTATCAAGG TCAGGGTGTT AAGACTATTT TACCTGCAGA AGCCAGTGCC AAGCTAGAGG	1980
TTCTGCTGGT TCCGGGCCTA GAACCGCATG ATGTTCTGGA AAAAATTCGG AAACAGCTAG	2040
ACAAAAATGG CTTTGATAAG GTAGAATTAT ACTATACCTT GGGAGAGATG AGCTATCGAA	2100
GCGATATGAG CGCACCAGCC ATTCTCAATG TGATCGAGTT GGCCAAGAAA TTCTATCCAC	2160
AGGGCGTTTC AGTCTTGCCG ACGACAGCGG GGACAGGACC TATGCATACG GTCTTTGATG	2220
CCCTAGAGGT ACCAATGGTT GCATTGCGTC TAGGAAATGC CAATAGCCGA GACCACGGTG	2280
GAGATGAAAA TGTGCGAATC GCTGATTATT ACACCCATAT CGAATTAGTA GAGGAGCTGA	2340
TTAGAAGCTA TGAGTAGAGA TATTATCAAG TTAGATCAGA TCGATGTGAC TTTTCACCL	2400
AAGAAGAGAA CCATCACAGC GGTTAAGGAT GTGACCATTC ACATCCAAGA AGGGGATATC	2460
TACGGAATCG TTGGATATTC TGGAGCAGGA AAATCAACCC TTGTACGGGT GATTAATCTC	2520
TTGCAAAAAC CATCTGCAGG GAAAATTACC ATTGACGACG ATGTGATTTT TGACGGCAAG	2580
GTGACCTTGA CGGCAGAGCA GTTGCGTCGT AAACGTCAAG ATATCGGAAT GATTTTCCAG	2640
CATTTTAACC TGATGAGCCA AAAGACAGCA GAGGAGAATG TAGCCTTTGC CCTTAAACAC	2700
TCTGAACTCA GCAAGGAAGA AAAGAAGGCT AAAGTAGCTA ACTTGTGGA CTTGGTTGGT	2760
TTGGCAGATC GTGCTGAAAA CTACCCTTCA CAACTATCTG GAGGGCAAAA ACAGCGTGTG	2820



1241

GCAATTGCGC GTGCCTTGGC CAATGATCCA AAAATCTTGA TTTCAGACGA GTCAACTTCT	2880
GCCCTTGATC CGAAGACAAC CAAGCAGATT TTGGCCTTGT TGCAAGATTT GAACCAAAAA	2940
TTAGGCTTGA CTGTGTCTT GATTACGCAT GAAATGCAGA TTGTCAAAGA CATTGCCAAC	3000
CGTGTTCAG TTATGCAGGA TGGGCATTTG ATTGAAGAGG GTAGTGTGCT TGAAATCTTC	3060
TCAAACCCTA AACAACCTTT GACTCAAGAC TTTATCTCAA CAGCTACAGG TATTGACGAA	3120
GCCATGGTCA AAATCGAGAA GCAAGAAATC GTGGAACACT TGTCTGAAAA CAGTCTCTTG	3180
GTGCAACTCA AGTACGCTGG AGCTTCAACA GACGAGCCAC TTTTGAATGA ATTGTACAAG	3240
CATTACCAAG TAATGGCTAA TATTCTCTAT GGAATATCG AAATTCTCGA TGGTACTCCT	3300
GTTGGAGAAT TGGTGGTGGT TTTGTCAGGT GAAAAGCAG CGTTGGCAGG TGCCCAAGAA	3360
GCCATTTCGTC AAGCAGGTGT ACAAATAAAA GTATTGAAGG GAGTACAGTA AGATGGAATC	3420
ATTGATTCAA ACCTATTTAC CAAATGTCTA TAAGATGGGT TGGGCTGGTC AGGCAGGCTG	3480
GGGAACGGCT ATCTACTTAA CTCTTTATAT GACAGTTCTT TCCTTCATTA TCGGAGGCTT	3540
CTTGGGGCTA GTGGCAGGTC TCTTTCTCGT CTGACAGCG CCAGGTGGTG TCTTGGAGAA	3600
TAAAGTCGTA TTCTGGATTT TAGACAAAAT TACCTCAATT TTTCGTGCGG TTCCCTTTAT	3660
CATCCTCTTG GCAATCTTGT CACCACTTTC TCACTTGATT GTTAAACAA GTATCGGGCC	3720
AAATGCAGCC CTGTCCAC TTTCTTTTGC AGTCTTTGCC TTCTGG	3766

(2) INFORMATION FOR SEQ ID NO: 226:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2520 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

TGTTGCTGAG TTAATCGGTA CGTTCATGTT TGTATTCGTC GGGACAGGAG CTGTTGTTTT	60
TGGAAATGGT CTTGATGGCC TTGGTCACCT TGGAAATCGCC TTTGCCTTTG GTTTGGCAAT	120
CGTGGTGGCA GCCTACTCAA TCGGAAGTGT TTCAGGTGCT CACTTGAACC CGGCTGTTTC	180
GATTGCTATG TTTGTAAACA AACGTTTGTG ATCTTCAGAA CTTGTAAACT ACATCCTTGG	240
TCAGGTTGTT GGAGCTTTCA TCGCTTCTGG CGCTGTCTTC TTCCTCTTGG CTAATCAGG	300
TATGTCAACT GCTAGTCTTG GTGAAAATGC CTTGGCAAAC GGTGTCAGT TCTTTGGTGG	360
TTTCTTGTGTT GAAGTCATCG CAACTTTCTT GTTTGTATTG GTTATCATGA CTGTGACTTC	420

1242

AGAAAGCAAG GGCAATGGCG CGATTGCTGG TTTGGTAATC GGTTCATCAT TGATGGCGAT	480
GATTCTTGTC GGATTGAAGA TTAAGTGGACT TTCAGTAAAC CCAGCTCGTA GCTTGGCACC	540
AGCTGTCTTG GTAGGCGGCG CAACTTTCAA CAAGTTTGGA TTTTCATCCT TGCACCAATC	600
GCTGGTGGAG TTCTGTCAGC CCTTGTGCA AAAAATTTCC TTGGAACAGA AGAATAATTG	660
AAACTCAAAA AGCCTTGCTC CTCATCTTGA GGAACAGGGC TTTTTCGTAT GATACTCTTC	720
GAAAACTCTCT TCAAACACG TCAGCTTCAT CTTGCCGTAG TATGGTTACT GACTTCGTCA	780
GTTCTATCCA CAACCTCAAA ACAGTGTGTT GATCTGACTT CGTCAGTTCT ATCTGCAACC	840
TCAAAACAGT GTTTTAAGCT GACTTCGTCA GTTCTATCTG CAACCTCAAA ACAGTGTGTT	900
AAGCTGACTT CGTCAGTTCT ATCTGCAACC TCAAAACAGT GTTTTAAGCT GACTTCGTCA	960
GTTCTATCTG CAACCTCAAA ACAGTGTGTT AAGCTGACTT CGTCAGTTCT ATCCACAACC	1020
TCAAAACAGT GTTTTGATCT GACTTCGTCA GTTCTATCCA CAACCTCAAA ACAGTGTGTT	1080
GATCTGACTT CGTCAGTTCT ATCCACAACC TCAAAACAGT GCTTTGAGCA ACCTGCGGCT	1140
AACTTCCTAG TTTGCTCTTT GATTTTCATT GAGTATGACT TTAGCGGTG TCAATTTCTT	1200
CTGGATAAAG GTCGTGTTGG AAGAGGCGTT GTTCTGCCAA GCCCTCATA TTAGTTCCTT	1260
GCTTACCGTA GTTGTAGTAG GGGTCGATTG AAATGCCACC GCGCGGAGTG AATTTTCCCC	1320
AGACTTCTAA ATAGCGAGGG TCTAGCAAGT TGACCAAGTC TTTCCCGATG GTGTTGATAC	1380
AGTTTTCGTG GAAATCTCCG TGGTTTCGGT AGCTAAATAG ATATAGTTTG AGGGATTTTG	1440
ACTCGACACA GAGCTTGTC GGAATGTAGG AAATATGAAT CGTCGCAAAG TCTGGCTGAG	1500
CAGTGATTTG TCCCAGCAGA GACATATCGA GGATATGGTG ACGAATGCC TGTTCCTTAG	1560
CGATTCTCT AGTAATTTGA ATTTGAGGT GATGACGTTG GCCGTAGGCA AAGGTGACAG	1620
CTTCGACTGT TTCATAGTGT TGCATGACCC AGAAAAGGCA GGTGTTGAA TCTTGACCAC	1680
CACTAAAGAC GACCAAGGCT AATTGACGTT TCATAGTACT CCTTCCAAAA TGGGAAATGT	1740
TCAGAGCAG CAAAAAGCTC CCATTAGGGA GCTAAAAAT ACCAAATCGA GGTTTTTTTA	1800
GCGATGGCAT ATCCCAAACA TCGTAATATT CTACTTATAT AGTAAATGA AATAAGAACA	1860
GGACAAATCG ATCAGGACAG TCAAATCGAT TTCTAACAAT GTTTTAGAAG TAGAGGTGTA	1920
CTATTCTAGT TTCAATCTAC TATAGTCTAG CATATTTTTT GAAAAATGGC AAAGGGCAAG	1980
AAAAAAGAGA CCAAAGAAAG TACTTGGTCT CTCGTTTGAT TAGCTCAATT CAGCAATGAT	2040
GGCCTTGATT TGTTCTGCTG TGTGAACACC TGCAACTGT TTGACAACTT GGCCGTCTTT	2100
TTTGAAGAGA AGAGTTGGAA TAGACATGAT TCCAAAAGCA CGAGCTGTGT TTGGATTTTC	2160
ATCAACGTCC ATTTTAACGA TTTTCAAGAC ATCTTCTGAA AGTTCTTCAG ACAATTTGTC	2220

1243

CAAGATTGGA CCTTGCATAC GACATGGACC ACACCAAGTT GCCCAGAAGT CTACTAAGAC	2280
CAAACCGTCT TTTGTTTCTT GTTCGAATGT TGCATCTGTA ATTGCTTTTG CCATTGTATT	2340
TCTCCTTTTT TTAGTTATAT TGGCTTAAAT CTTGTTTCAT GAGATAGAAG AAGATATCTC	2400
CATAAGTCCC ATGGTAGTCC AAATTATGAC CCTTGTAAGT TAATTTTGG ACAGGGTAGT	2460
AKKCTGCGAC GCCGATAAGG CAAGCTTGTT GCGAACGTTT AAAGTCTTCA TAAGACTCGG	2520

(2) INFORMATION FOR SEQ ID NO: 227:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5278 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

ACTCAGTTAG ATTTTGTTTT CAAAAACAAC GAAGAAAAAG ACCATGTTGC TCTACTTGGA	60
AGAATTGGCT CCGAACGTGT TTATCGATAT ATTAATAAAA AATATTTAGA TTTACCGGAA	120
ACATTCGAAA ATTATAATGT TTTTGTACCA GAAGCTAATG GAAGTGGTGC CTTAGGTGAA	180
GTCTTATCAA CACCCCTAAT CGGGGAACCC CTAATCGGGC ATACAGATAC TTTTTATCT	240
ATTGGTAATT TAAAAACAAA ATTTGAAGCC GATGCTTGTA TTAAATTTAT TAAAACTAAA	300
TTCGCTAGAG TATTATTAGG TGTTTTGAAA GTTACTCAGC ATAATTCAGC CAAAACCTGG	360
TATTACGTCC CCCTCCAAGA CTTTACGGTC AATTCGGACA TTGATTGGAC ACAATCAGTG	420
ACTGATATTG ACCGCCAGCT TGATCAAAAA TATGACTTTT CCCCTGAAGA AATTGCCTTT	480
ATTGAGAATC ATGTAAGGGA GATGGATTAG AAAAGTATTT TTATTTGACA AATAGTGCTC	540
AATGATCTAA AATGACTATA TAGGATTAGG TCAGGAAGCA TACGATGCCC TGACCCTTTT	600
TGTAATTATG AGATGAGAAA GTCATTTGTT AGATAAATTG ACTCGTTAGC AAACGTTCAA	660
AAAAGGAAAA CTTATGCCAG TAGAAATTAA AACCCTAAA GAAATTCATC CTAAATCTA	720
TGCCTACACC ACACCGACAG TAACCAAGTAA TGAAGGCTGG ATTAAGATTG GGTATACAGA	780
ACGTGATGTC ACACAACGTA TCAAGGAGCA AACGCATACA GTCATATAG CTACAGATGT	840
CTTATGGACT GGTGATGCAG CTTATACAGA AGAGCCTGAT AAGGGGAAAA CTTTCAAGGA	900
CCATGATTTC CACCATTTC TTTCTTTCCA TGATGTAGAA CGTCGTCCCA AGACGGAATG	960
GTCTATTTT AATGGAATC CTGAAAAATC AAAAAATCTT TTTGATAAGT TTGTTACAGA	1020
TGATTTGTCT GGTATCAGC CTGAAAAAGG ACAGGACTAT ACTCTGCGAC AAGAGCAAGA	1080

1244

AGAAGCAGTT GCTAAGACAT TAGCTTATTT CCAAGAACAT GCTGGAGGCA AGTTTCTCTG	1140
GAATGCCAAG CCACGCTTTG GTAAAACCTT GTCTACCTAT GACCTAGCTC GACGGATGGA	1200
AGCTGTCAAT GTCCTAATTG TAACAAACCG CCCTGCCATT GCTAACTCAT GGTATGATGA	1260
TTTTGAAACA TTCATAGCAG GTCAAACGAC TTACAAGTTT GTTCTGAAT CAGATAGCCT	1320
TAAGAGTCGT CCAATCTTGT CACGACAAGA ATTTCTTGGT ATTTTAGCTG ACGATGTAAG	1380
ACAACTTGCT TTTATCAGTC TCCAAGACTT GAAAGGATCT GTTATTTAG GTGGAGAGCA	1440
CGATAAACTC AAATGGGTAA CTGATCTGCA TTGGGACTTG TTGGTTATTG ACGAGGCTCA	1500
TGAAGGAGTT GATACCTTCA AGACTGACCA AGCCTTTAAT AAGATTGAC GAAATTTTAC	1560
TCTGCATTTG TCAGGTACAT CATTTAAAGC ATTGGCTAAA GGAGATTTTA CAGAGGAACA	1620
AATCTACAAC TGGTCTTATG CTGATGAGCA GGCTGCTAAG TATTCGTGGT CTCTTGAGCA	1680
AGAAGAGGAA AATCCTTATG AAAGCTTGCC TCAGTTGAAT CTCTTTACCT ATCAAATGTC	1740
TCAGATGATT GGCGAAAAGT TAGAAAAAGG CGCTCAGATC GATGGTGAAT ATATTGACTA	1800
TGTTTTTGAC TTAAGTGAAT TTTTCGCTAC AGATGATAAA GGGAAATTTA TTCATGAGCA	1860
TGATGTCAGA AATTGGTTAG ATACTCTATC AAGCAATGAA AAATATCCAT TTTCAACCAA	1920
AGAACTCCGT AATGAACTCA AGCATACTTT TTGGCTTTTA GAACGTGTCG CTTCGGCCAA	1980
AGCATTAAAA GCCCTACTAG AAGAACACCC AATCTATGAA AACTATGAGA TCGTTCTAGC	2040
TGCTGGTGAC GGACGTATGT CCGAAGAAGA CGATAAAGTC AAACCTCAAAT CCTTGGACTT	2100
GGTTAGAAAA GCGATAGCAG AGAATGACAA AACCATTACC CTATCCGTTG GTCAGCTGAC	2160
GACAGGTGTC ACTATCCCTG AATGGACAGG TGTATTGATG TTATCAAATT TGAAATCACC	2220
AGCTCTTTAT ATGCAGGCCG CCTTCCGTGC TCAAATCCT TACTCATGGA GCGATAACAA	2280
AGGAAATCAC TTTGCGAAAAG AAAGAGCCTA TGTATTTGAC TTGCGCCGG AAAGAACCTT	2340
GATTCTCTTT GATGAGTTTG CCAACAACCT ATTGCTTGTA ACTGCAGCTG GTAGAGGAAC	2400
TTCAAGTACA CGCGAAGAAA ATATTAGAGA ATTATTAAC TTCTTTCCAA TTATTGCCAA	2460
AGACCGTGCT GGTAAGATGG TTGAAATTGA TGCAAAGGCA GTTCTAACCA CTCCTCGCCA	2520
GATAAAAGCT AGAGAAGTTC TTAACGAGG TTTTATGTCC AATCTCTTAT TTGATAATAT	2580
TAGTGGTATT TTCCAAGCAA GTCAAACAGT TTTAGATATT TTAATGAGC TGCCAGTTGA	2640
AAAGGAAGGG AAGGTACAAG ATAGTTCTGA TTTATTAGAT TTTTCAGATG TTACAGTCGA	2700
TGATGAGGGA AATGCAGTAG TAGACCATGA AATTGTAGTT AATCAGCAAA TGCGACTTTT	2760
TGGTGAAGAA GTTATGGAC TTGGTGAATC TGTGCTGAG TTAGTCACAA AAGATGAGGA	2820
ACGAACTCAA AAACAGCTGG TCAATGACTT GAGTAAGACC GTTCTTCAG TGATTGTAGA	2880

1245

GGAATTGAAA GCAGATTATT CTCTAAAAAC AAGGGAAACT GAGCAAATTA AGAAACAAAT	2940
TACAGCAACA CTTGAGAATG AAATTCGAAA AAATGATATC GAAAGAAAAA TTTCTGAAGC	3000
TCATATCAAG CAAGAGTTGC AACAGCAGCT CAAAGAAGCA AATGATAAAG CGCAAAAAGA	3060
TAAGATTCAA GAAGATTTGG AAAACGTTT AGAAGAAAAT AAATCATTTC ATAAAGAAAA	3120
ACTAGAACAA AACTCAAAA AAGAAGTGGG AAAAATGCCT GAGAAATTTA TCGAACAGGT	3180
TGAGATAAAA CGTGTGGAAC AGTTGAAACA ATCAGCTCAA GATGAAATTC GTGACCATT	3240
ACGAGGGTTT GCAAGAACAA TTCCAAGTTT TATTATGGCT TACGGTGATC AAATCTAAC	3300
ACTTGATAAT TTTGATGCCT TTGTTCTGA ACATGTTTTT TATGAAGTAA CAGGGATTAC	3360
GATTGATCAG TTTAGATATT TGCGAGATGG TGGGCAGGAT TTTGCAGGGC ATCTCTTTGA	3420
TAAAGCAACA TTTGACGAAG CTATTCAAGA ATTTCTTCGC AAGAAAAAGG AGTTGGCGGA	3480
TTATTTTAAA GATCAAAAAG AAGACATTTT TGACTATATT CCACCGCAGA AGACCAACCA	3540
AATTTTCACT CCTAACGAG TGGTGAAAAG GATGGTAGAT GATTTGGAAA AGGAAAATCC	3600
AGGGATTTTT GATGATCCAT CTAAGACTTT TATTGATTTA TATATGAAGT CAGGCCTCTA	3660
TATTGCAGAA CTTGTGAAGC GGTATATATA TAGCAATGGC TTGAAAGAGG CCTTTCCAAA	3720
TCCTGAAGAA CGCTTAAAC ATATTTTGGG AAAGCAAGTT TATGGATTG CTCCGTCTGA	3780
GATTATCTAT AACATTTCCA CTAATTTTAT ATTTGGCAAT CTTTCTAAAG ATATCAGTAG	3840
GAAGAATTTT GTTTTAGCAG ATACCATTCC AGCGGCTAAA GAAGGGAGCA TTCAAAAGTT	3900
GTTTGATTCC TATTTTGAAA ATAATTAAAA AGAAGGCCGA GTCAAAATTC TTTGAAATCA	3960
GAAAAACGC ATAATATTGA GTGCTTTTGT ACTGCCCCC AAAAGTTAGA CAGAAAAAAT	4020
CTAACTTTTG GGGGGCAGTT CAGACAATCC TTGGTATTAT GCGTTTTATT GTGGGAAGAT	4080
GTATAATGGA TTGAAATAAG ATATGAACAA ATCAATTAGG AATTTAAAGC ATTTTATAAC	4140
AACGTTTTAG AGTAATGGG GGCTATTTCA ACTTCAACCT ACTATAATAC AGAAAAAAC	4200
AACTCCCTGA TAATTCAAGG AGTTGTCTAT AGTTAAATTA GTTTTTAGAA GCTTCTTGGA	4260
ATTCTGGGTT TTTCCATGCT TCGTCAATGA TAGCTTGTA TTTCTTAGCA GATGCTTGCA	4320
TTTTTTGAGT TTCTGCGTCG TTCAATGGGA TATTTACTGG ACGAACGATA CCATGTGCAC	4380
CAACAACAGC TGGTTGACCG ATAAAGACAT TCTCAACTCC GTATTGACCT TCTTGAATA	4440
CTGAAAGTGG AAGTACTGCG TTTTCATCGT CAAGGATTGC TTTAGTGATA CGAGCAAGGG	4500
CTACTGCGAT ACCGTAGTAT GTTGCACCTT TTTTGTGAT GATTGTGTAG GCTGCATCAC	4560
GAACACCTTC GAACAATTCA ATCAATTCAG CTTCTTGAAC ATTTTGAGTG TCTTTAAGGA	4620

1246

ATTCTTCAAG GTTTACACCA GCGATGTTAG CGTGTGACCA AACAGCGAAC TCAGAGTCAC	4680
CGTGTTCACC CATGATGTAG GCGTGCACTG AACGAGCATC CACATCCAAT TTTTCAGCAA	4740
GTGCTTGACG GAAACGAGCT GAGTCAAGTG AAGTACCTGA ACCGATAACG CGTTCTTTAG	4800
GGAAACCAGA GAATTTCCAA GTTGAGTAAG TCAAAACGTC AACTGGGTTA GCAGCAACAA	4860
GGAAGATACC TTTGAAACCA GATTCAACAA CTTGAGTTAC GATTGATTG TTGATAGCAA	4920
GGTTTTTACC TACAAGGTCA AGACGAGTTT CACCTGGTTT TTGAGGTGCA CCTGCAGTGA	4980
TCACAACAAG GTCAGCGTCT GCACAGTCAG AGTATTGAGC TGCATAGATT TTTTATAGTG	5040
AAGTGAAGGC AAGGGCGTGA CTAAGGTCAA GCGCATCACC AACAGCTTTT TCATGCAATT	5100
GTGGAATTC GATAATTCCA AGCTCTTG TG CAATTCCTTG GTTAACAAGT GCAAAAGCGT	5160
AAGATGAACC TACAGCACCA TCACCGACAA GGATACTTT TTTGTGTGT TTAGTTGAAG	5220
TCATTGTTTT AAACATCTCC TTAATTTTAT TAGGGGATTT TCCCTAGACA ACTTCATT	5278

(2) INFORMATION FOR SEQ ID NO: 228:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1941 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

ATAAGGAATC TCTAAAAAAT TTAAAGGAGA ATCTAGCAAA TGGATTTCAC ATGGGCACTG	60
AAGTATGCCA CTGAATTTTT GGGAACTGCC ATTTTGATCA TTCTTGGGAA TGGTGCAATT	120
GCCAACGTTG AACTTAAAGG TACGAAAGGT CACCAAAGTG GCTGGATCGT CATCGCTGTT	180
GGTTATGGTA TGGGGGTTAT GATCCCAGCC TTGATGTTTG GTAACGTATC TGGGAATCAC	240
ATCAACCCTG CTTTCACTCT AGGGCTTGCA GTTAGCGGTC TTTTCCCTTG GGCACAAGTG	300
GTACCTTACA TTATCGCGCA AGTCTTGGGG GCTATCTTTG GCCAAGCCTT AGTTGTGGCA	360
ACATACCGTC CATTCTACTT GAAAAGTGAA AACCCTAATA ACATCTTGGG AACTTTCTCA	420
ACTATTTCAA GTATTGACCA TGGTACAAAA GAAAGTCGCT ATGCAGCAAC TGTCAATGGT	480
TTGATTAAATG AGTTTGTGG TTCAATTGTT TGTTCCTTTG CAGCTCTTGG TTTGACTAAA	540
AACTTCTTTG GTGCTGAAGT GCTTCAATTC ATGAAACAAA AGGCAACAGA AGCAGGACAA	600
ACAGTTGATT TTTCTGACTT GGCTATTAAG GCACAGGTGG CTCCACACAC TGCTTCAGGA	660
CTTTCTGTGG CTCACCTGGC ACTTGGATTC CTCGTTATGG CTTTGTTAAC ATCACTTGGA	720
GGACCTACAG GACCTGCCTT GAACCCAGCC CGTGACTTGG GACCACGTCT CCTTCATGCT	780

1247

TTCCTTCCCA AATCAGTTCT TGGTGAGCAT AAAGGCCGATT CAAAATGGTG GTATTCTTGG	840
GTACCACTAG TAGCACCTAT CGCAGCAGCA ATTGCGGCAG TAGCTGTATT CAAATTCCTT	900
TATCTCTAAG AAATAGCTCC TTAAACATTT GAGTGAGCAC CATCTATAAG TAAGAGAGGA	960
TCAGACTGGk TCTCTCTTTT kGATTTTtAG GGAAATGAAA GAAcTCTAAA CAAACTCCTC	1020
TCCAGCAGTG GTTTAGAAGT CTCAGTGGGC TATTCCAGCT TCAATGGACT ATAGTAGGTT	1080
GCAGTTGAAA TAATAGACCC TTGTTTCTAA AACATTGTGA GAAATTGGTT TGAATTCTCC	1140
AATCAAATG TGCAGTTTTC ATTCTACTAT ATATTATCGG AATATTATCG GAGATGGGTT	1200
CCCTATCTTG TAAGTCTGCT TTATAGTGGG TTGAAGTTGG AATAGTCCTC CCTTCTTTCT	1260
CAAACATTGT GAGGAATTGA TTTACCTTCC TCAACAAAAT GTTCAGTTTC TATTTCAATT	1320
TACTATAAAA TAAGCGATTA GGGGGGCTAT TCTTCGACCT ACATTGACTC TGCTGAGTCC	1380
TATGATTGTT ATCGTTTAT CTGCAATTTT AACTCAATG AAAATCAAAG GGCAAACATA	1440
GAAGCTAGCC GCAGGTGTT CAAAACACAG TTTTGAGGTT GTATAGTAGA TTGAAACTAG	1500
AATAGTACAC ATCTACTTCT AAAACATTGT TAGAAATCGA TTTGACTGTC CTGAACGATT	1560
TGCCCTATTC TTGTTTCATT TTAATATATA AACCAGAGAC TGTTCACATT TTCAGCAAGT	1620
GAGTGGATGG ATAATGCTGA AAACCTCTTG AAGGATAAGT CTATTTAGTA CTTTCTATTA	1680
ATTAGTTAAA TTTTACCAA GAATAATTCA CAAAACGTT GTAAACACT TGCAATTTAG	1740
CTGAAATTTG ATAAAATAGT AAGGAAAGTT AGACTGTATT GCCTACTGTC TATCTATAAA	1800
ATATATTTTA TTGGAGGCTT TTAATCAAAT GGCAAAAGAA AAATACGATC GTAGTAAACC	1860
ACACGTTAAC ATTGGTACTA TCGGACACGT TGACCACGGT AAAACTACCC TAACTGCAGC	1920
TATCACAACT GTTTTGGCAC G	1941

## (2) INFORMATION FOR SEQ ID NO: 229:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 755 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

ATTTGAAGAA ATTGAAGAAA TCGTAGCCCC TACAGATGGT GAATTTTGG GGGAAGTTTT	60
ACTTGGAAC GGGGTAGTTC TCTTAATTGG AGTAGCCTGT TGTAAAAAG ATAGGGAGTG	120
ATAATCATGC AAGATAACTT TTTATTGAG GAAATTGAAG AAATTTCACT ACCAGTTAAT	180

1248

GATTTTTCAG CTGGACTTGC AACAGGTATC GGATTTGGTT TAGCAATCCT TGCTCTTGCT	240
GGTTGTTGAA GTTTGTTTCAT TTTACTAACAT CAAGCTTTTT CAATTTTCATT TTAGACAGTC	300
ATTTAAATTT TCCGTATTAG TCTTGCAGCA AGAGATTAAT AGAATTAGTC ATTATTTTAT	360
TGATTGCGGA CTGAGGGACT AGAGTATGTT TTTACTTAACC CCTCTTTTAT TTATTAAAGG	420
TTAGGTTTGT TATGAGAATT GTTGATAAGA TTAAGATATT ACCTACTCCT TATGAGGGAC	480
ACTATCATTT ATATATACCA TCCAGTAAGA AACATGTATT AGTTGGGAAA CAGGAAAAAA	540
ATGGTTAGAG CAACTAATAG GTCAAGAATT TACCATATCG GACTTATTAG TGTTAGTAGG	600
GAAGAAATAT TTTTAAATA TCTTGGGACT TTAATATAAC ATTATCTGAA AAATTAAACT	660
ATAAAAGATT TAATAAGAAT TTTGAAAAAA TCCTATCTTG TTGTCATTAT ATTTGCAACG	720
ATACATGAAA TTAGTCATGC AATAATTGCT AATAA	755

## (2) INFORMATION FOR SEQ ID NO: 230:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1483 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

CCAGAAAAAC CGTAGTGGAG CTCGTGGAAC AGTGGAATTG ATTTTCCAAA AAGAATACAA	60
TAAATTTTCA AGTATCTCAA AGAGGGAGGC ATAAGATGTC AGATGCATTT ACAGATGTAG	120
CCAAGATGAA AAAAATCAAA GAAGAAATCA AGGCACATGA GGGACAAGTC GTAGAAATGA	180
CTTTGGAGAA TGGTCGTAAG CGCCAAAAAA ATAGATTGGG TAAGCTAATT GAAGTTTATC	240
CATCTCTATT TATTGTGGAG TTTGGGGATG TGAAGGAGA TAAACAAGTT AATGTTTACG	300
TTGAATCCTT TACTTACTCA GATATTCTTA CAGAAAAGAA TTTGATTCAT TATCTTGACT	360
AAAGTGAGAA ATTTTCTCAC TTTTCTTTT TTCTCCGAAT AATTTAGGTG AAGGCAATCA	420
TCGCTTTATA TTATTTTTC AAGAGGAAGA ATGAAAATTT TACCGTTTAT AGCAAGAGGA	480
ACAAGTTATT ACTTGAAGAT GTCAGTTAAA AAGCTTGTTT CTTTTTTAGT AGTAGGATTG	540
ATGCTAGCAG CTGGTGATAG TGTCTATGCC TATTCCAGAG GAAATGGATC GATTGCGCGT	600
GGGGATGATT ATCCTGCTTA TTATAAAAAT GGGAGCCAGG AGATTGATCA GTGGCGCATG	660
TATCTCGTC AGTGTACTTC TTTGTAGCC TTTCGTTGA GTAATGTCAA TGGTTTGTAA	720
ATTCGGGCAG CTTATGGAAT TCGGAATGAA TGGGGACATC GTGCTCCTCG GGAAGGTTAT	780
CGTGTAGATA ATACCCGAC GATTGGTTCC ATTACTTGGT TACTGCAGG AACTTATGGT	840



1249

CATGTTGCCT GGGTGTCAAA TGTAATGGGA GATCAGATTG AGATTGAGGA ATATAACTAT	900
G GTTATACAG AATCCTATAA TAAACGAGTT ATAAAAGCAA ACACGATGAC AGGATTATT	960
CATTTTAAAG ATTTGGATGG TGGCAGTGTT GGGAATAGTC AATCCTCAAC TTCAACAGGC	1020
GGAACCTATT ATTTTAAGAC CAAGTCTGCT ATTAAACTG AACCTCTAGC TAGCGGAAC	1080
GTGATTGATT ACTATTATCC TGGGGAGAAG GTTCATTATG ATCAGATACT TGAAAAAGAC	1140
GGCTATAAGT GGTGAGTTA TACTGCCTAT AATGGAAGCT ATCGTTATGT TCAATTGGAG	1200
GCTGTGAATA AAAATCCTCT AGGTAACTCT GTTCTTTCTT CAACAGGTGG AACTCATTAT	1260
TTTAAGACCA AGTCTGCTAT CAAAAGTAA CCCCTAGTTA GTGCAACTGT GATTGATTAC	1320
TATTATCCTG GAGAGAAGGT TCATTATGAT CAAATTCTCG AAAAAGACGG CTACAAGTGG	1380
TTGAGTTATA CGGCTTATAA CGGAAGTCGT CGCTATATAC AGCTAGAGGG AGTGACTTCT	1440
TCACAAAATT ATCAGAATCA ATCAGGAAAC ATCTCTAGCT ATG	1483

(2) INFORMATION FOR SEQ ID NO: 231:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1027 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

CCCGGAAAAC AAGTTAAAGT TGAAGTTGGT CAAGCAGTTT ACGTTGAAAA ATTGAACGTT	60
GAAGCTGGTC AAGAAGTTAC TTTTAACGAA TTGTTCTTGT TGGTGGTGAA AACACTGTTG	120
TCGGAACTCC ACTTGTTGCT GGAGCTACTG TAGTTGGAAC TGTTGAAAAA CAAGGAAAAC	180
AAAAGAAAGT GGTACTTAC AAGTACAAAC CTAAAAAAGG TAGCCACCGT AAACAAGGTC	240
ACCGTCAACC ATATACAAAA GTTGTCATCA ACGCAATCAA CGCTTAATTT TAAGGAGAAC	300
ACATGATACA GGCAGTCTTT GAGAGAGCCG AAGATGGCGA GCTGAGGAGT GCGGAAATTA	360
CTGGACACGC CGAGAGTGGC GAATACGGCT TAGATGTCGT GTGTGCATCG GTTTCTACGC	420
TTGCCATTAA CTTTATCAAT TCTATTGAGA AATTGTCAGG CTATGAACCA ATCCTAGAAT	480
TAAACGAAGA TGAAGGTGGC TATCTGATGG TTGAAATACC AAAAGATCTT CCTTCACACC	540
AGAGAGAAAT GACCCAGTTA TTCTTTGAAT CATTTTCTT AGGTATGGCA AACTTATCGG	600
AGAACTATTC TGAGTTCGTC CAAACCAGAG TTATCACAGA AACTAACAC GGAGGAAAAC	660
ATTATGTTAA AAATGACTCT TAACAACTTG CAACTTTTCG CCCACAAAAA AGGTGGAGGT	720

1250

TCTACATCAA	ACGGACGTGA	TTCACAAGCA	AAACGTCTTG	GAGCTAAAGC	AGCTGACGGA	780
CAAACGTGAA	CAGGTGGATC	AATCCTTTAC	CGTCAACGTG	GTACACACAT	CTATCCAGGT	840
GTAAACGTTG	GTCTGGTGG	AGATGATACT	TTGTTGCTA	AAGTTGAAGG	CGTAGTACGC	900
TTTGAACGTA	AAGGACGCGA	TAAAAAACAA	GTGTCTGTTT	ACCCAATCGC	TAAATAAAAA	960
GGTCCATTGA	ACCTTTTATC	CCGAACCTTG	AAATGTAGAG	GTGAGGAAGC	TAGAAACAGC	1020
TTAAAAAT						1027

## (2) INFORMATION FOR SEQ ID NO: 232:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1990 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CGGTTCAAAT	GGTGCAGGTA	AATCTACGTT	AATTAATTCT	ATTGTAGGTT	TTCAAGAGAT	60
TTATTTAGGA	GAAATAGAGT	ATTGTGATAA	AGATTTGATA	GTTAGTTCTC	AACCTTTTGC	120
TCATTTAGGC	TTTACTCCTC	AAACCACAGT	AATTGATTTT	TATACTACTG	TGAAGGACAA	180
TGTAATATTG	GGGCTGAACC	TTGCTGGAAA	GTTTGGGAAA	AATGCTGAGA	AGTTGTGTCA	240
AATAGCCTTA	GAAATTGTTG	GGTTAGCTGA	TAAAAAAAAT	AATTGGTAG	AAACATTGTC	300
AGGTGGACAA	CTGCAACGCG	TCCAGATTGC	TAGAGCAATA	GCTCATAATC	CAGATTTTTC	360
TATTTTAGAT	GAACCTACCG	TTGGTTTAGA	TACTGAATCT	GCCGAAAAAT	TTTAAATGTA	420
TTTAAAAGAT	AAGAGTTTGG	AAGGAAAAAC	TATTATCATA	TCCTCACATG	ACATAAATCT	480
ACTCGAAAAG	TTTTGTAAAA	AAATACTTTT	TTTACAAAAT	GGCTCCATAT	CATTMTTTTG	540
TGATATGCGT	GACTTTGTAG	ATAATTCAAC	TATCAAATTA	AATTTTTCAT	TGCAGAATAG	600
AATTTCTAGA	TATCAAATTG	AATTTTGTGA	AAATTTTAGA	TTTAAAGTTC	ACATCGAAGA	660
TAATGATAGT	TTTACAATAG	AAGTCCCTAT	AGAAGAAAAG	ATCTTAGATG	TTATCAATGA	720
GGTAGGAAAA	GCATGTGAAA	TTAAAACTT	TTCAACAAGT	AAATTAACCT	TACAAGAAAG	780
TTATTTGCAA	AGAATAGGAG	GAGAAAAATG	AAGGCTGATC	AATTAAGGCA	CAAATCGGAC	840
TTAGGTTTAA	GAGGTCTAGC	GATTATTGCT	AAAAATGAGA	TTATTGCTTT	TTTTAGAAGT	900
AAAGGTTTAA	TTATTTCTCA	GTTTCTACAA	CCAATCTTAT	ATGTTGTTTT	TATAATAATA	960
GGATTAAATT	CTTCGATAAA	GAACATTCAG	TTTAATGATA	TAAAAACCTC	TTATGCAGAA	1020
TATACAATCA	TTGGTGTTAT	AGCTTTATTG	ATAATCGGGC	AGATGACTCA	AGTTATTTAT	1080

1251

AGGGTGACAA TAGATAAAAA ATATGGGCTA CTTGCTCTTA AGTTATGCAG TGGAGTTCGT	1140
CCTTTATATT ATATTTTAGG GATGAGTATC TATTCTATAT TAGGGTTGAT AGTTCAAGAA	1200
ATTATTATAT ATATAATTAC GTTAGCGTTT GAGATAAATA TCGCAATGGA TAGATTTTTT	1260
TATACAGTTT TGTATCTAT TGTGTTTTTA TTATTTTGGG ACTCCCTTGC AATTTTACTT	1320
ACAATGTTTA TCAATGATTA CAGAAGACGT GATATTGTAA TACGTTTTGT ACTAACACCG	1380
CTTGGTTTTA CAGCTCCTGT TTTCTACTTA ATAGATTCTG CTCCTAGTAT TGTGAGATGG	1440
ATTGGTCAGT TAAATCCCTT AACTTATCAA TTAATATTT TGAGAACTT TTATTTTAAA	1500
AATTCAACAA CTTTGAATT AGTTTCTTA TTGTTAACAT CATTACTTGT CCTTATATCT	1560
GTATCTTTTA TTATACAAA GATAAAATTG ATACTGATAG AAAGATAAAA GTTGGGTCAT	1620
CCAACTTTTT TGTGTCTCC CGAAAACCAC TAGCTATGCT AGTGGTTCCA TAGAGCTTTT	1680
AGCGTGGTAA CAAAAGAAC CTCCTAAAAT GATAAGATAG AAGTGGTTTC TCCGCCACTA	1740
CAACATATCA TACAGGAGGT ACCTCATGAG AGAGGATAAT CAAAGTTTAT CACATACCAC	1800
ATGGAATTGT AAATATCATA TTGTTTTTGC ACCCAAATAT CGTCGTCAA TCATTTATGG	1860
CAGATACAAA GCTAGTATCG GAAGAATCAT ACGTGACTTA TGTGAGCGTA AGGGTGTAAT	1920
AATCCATGAA CGGAATGCTT GTTCAGACCA TATTCACATG CTTATCAGTA TTCCTCCGAA	1980
ACTTAGTGTT	1990

(2) INFORMATION FOR SEQ ID NO: 233:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4766 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

GAACATATATT GCATATATTT CTAGCAATGA TCATGGCGAA TCTTGGTCTG CACCAACTTT	60
ATTACCTCCT ATAATGGGAC TTAATCGGAA TGCGCCATAT TTAGGTCCTG GACGTGGAAT	120
CATTGAAAGC TCAACTGGAC GTATTCTTAT TCCGTCTTAC ACTGGTAAAG AGTCTGCGTT	180
CATTTATAGT GACGATAATG GAGCATCTTG GAAAGTTAAA GTAGTGCCAC TTCCTTCTAG	240
TTGGTCAGCA GAAGCACAAT TTGTAGAATT GAGTCCAGGA GTAATTCAAG CATATATGCG	300
TACAAATAAT GGTAAAATTG CATATTTAAC AAGTAAAGAC GCAGGTACTA CTTGGAGTGC	360
ACCGGAATAT TTGAAATTG TTCAAATCC AAGTTATGGA ACACAATTAT CAATCATCAA	420

1252

TTATAGCCAA	TTGATTGATG	GTAAAAAGGC	TGTCATTTTA	AGTACTCCAA	ACTCCACAAA	480
TGGTCGTAAA	CACGGACAAA	TTTGGATTGG	TCTAATTAAT	GATGATAATA	CAATTGATTG	540
GCGTTATCAT	CACGACGTTG	ATTATAGTAA	CTATGGATAC	TCATATTCAA	CATTGACAGA	600
GTTACCAAAT	CATGAAATTG	GATTGATGTT	TGAAAAATTT	GATTCATGGT	CTCGTAATGA	660
ACTTCATATG	AAAAATGTTG	TACCATATAT	AACATTTAAG	ATTGAAGATC	TGAAAAAGAA	720
TTAAAGCTGA	AATTTGAAAA	TATATAAAAA	GAGGATAAAA	ATTATGGTAA	ATTACGGTAT	780
TGTTGGAGCT	GGATATTTTG	GAGCTGATTT	AGCTCGCTCA	ATGAACAAAA	TTGAAGATGC	840
AAAAGTGGTT	GCGGTATTTG	ACCCAAATCA	TGGAGAAGAA	GTTGCTCAAG	AGTTGGGATC	900
AGATGTTTGT	GCAAGTTTAG	ATGAACTTGT	AGCACGTGAA	GATATTGATT	GTCTGATCGT	960
AGCTTCACCT	AGCTACCTTC	ACCGTGAACC	AGTTGTGAAA	GCTGCTCAAC	ATGGCAAACA	1020
CGTATTTTGT	GAAAAGCCAA	TTGCATTGTC	TTATGAAGAT	TGTAAAGCCA	TGGTTGACGC	1080
ATGTAAAGAA	AATAATGTCA	TCTTTATGGC	TGGTCACATC	ATGAACTTCT	TTAACGGTGT	1140
ACACCATGCT	AAACAATTGA	TTACTCAAGG	TAAAAATCGGT	AAAGTTCTTT	ATTGCCATGC	1200
TGCTCGTACA	GGTTGGGAAG	AACAACAACC	AATGTATCA	TGGAAGAAAC	TTCTGTTCTCA	1260
ATCTGGAGGA	CATTTGTACC	ACCATATTCA	TGAATTAGAT	TGCATTCACT	TTATCATGGG	1320
AGGACTTCCT	GAAAAAGCGA	CAATGGTAGG	AGGCAATGTA	TATCATAAAG	GTGAAAACTT	1380
TGGTGATGAA	GATGATATGC	TCATTGTAAA	CTTAGAATAC	TCTGATGATC	GTTATGCTGT	1440
TTTGAATAT	GGTAATGCTT	TCCGTTGGGG	TGAACACTAC	GTCTTGATTG	AAGGAACTGA	1500
AGGAGCTATC	AAACTTGACT	TGTTCAATAC	TGGCGGTACT	CTTCGTGTTA	AAGGTGAAGG	1560
AGAATCACAC	TTCTTAGTTC	ATGAAACTCA	AGAGGAAGAT	GATGATCGTA	CAGCTATCTA	1620
TACCGGTCGT	GGTATGGATG	GAGCAATTGC	GTACGGTAAA	CCAGGAGTAC	GTTGCCCATT	1680
ATGGTTGCAA	ACATGTATTG	ATAAAGAAAT	GGAATATCTA	CATGACATCA	TTAAAGGTGG	1740
AGAAATTACA	GAAGAATTG	AAAAACTTCT	CAATGGTGTA	GCTGCTTTAG	AATCAATCCT	1800
TACCGCTGAT	GCATGTACTT	TATCAGTTAA	AGAAGATCGA	AAAGTAAGTC	TTTCAGAAAT	1860
CACAAATGCT	TAACTTTTGT	AAAACAGAAT	AGTAAATTCT	TGTCATTATA	TAATTTCTAA	1920
AGTTCTGTGA	TACAACTCAT	TGAATAAAGA	AATAGAGATG	GGACTGGGAT	AATGCCCAGT	1980
CCCATTTTTT	ATCAAAAAGT	AATGAGATCA	AAAATGTGGG	AGTGTGAAA	TGAAGATTAT	2040
AGGTATCGAT	ATTGGCGGAA	CAACAATTAA	GGCAGATTTA	TACGATGAGT	TTGGAACGAG	2100
TTTGAATCAT	TTCAAAGAGA	TAGAAACAAT	TATTGACTAT	GATTTGGGAA	CGAATCAGAT	2160
ATTAAATCAG	GTCTGTGATT	TAATGGTGA	GTATACTTTA	AATCATTCAA	TTGATGGTGT	2220

1253

TGGGATTTC	ACTGCTGGAG	TTGTTAATGC	TAATACTGGA	GAAATCATCT	ATGCAGGCTA	2280
TACAATACCA	GGGTATATCG	GAGTAAACTT	TACTGCCGAA	ATAGAAAAAC	GTTTTGGGTT	2340
GTATACTTTT	GTGAAAAATG	ATGTTAATTG	TGCTGCATTA	GGTGAATTGT	GGAAGGGACA	2400
AGCCAAAGAT	AAGAAAAATG	TAGTAATGGT	TACTATTGGA	ACAGGTATAG	GAGGCAGTAT	2460
TATTGTCAAC	GGACAAATTG	TTAACGGATT	TAACTATACT	GCTGGTGAAG	TAGGTTATAT	2520
TCCTGTAGGT	AATTCGGATT	GGCAAAGTAA	AGCCTCAACA	ACCGCATTGA	TTCATTTATA	2580
TCAAAAAAG	AGCTTGAAAA	CTAATCAAAC	TGGACGTACT	TTCTTCACTG	ATTTAAGATC	2640
TGGAGATAAA	GTTGCTGAAG	AAACTTTTGA	AATTTTGTGA	GAAAACTAA	CAAAAGGTTT	2700
ATTAACGATT	TCTTATCTAC	TTAATCCAGA	AATTCCTATA	TTAGGAGGTG	GGATTCTGGA	2760
TAGTAAGGAT	ATTTTGTTAC	CTGAAATTCA	AAGTTCTTTA	GCTAAAAATG	CAATGGATAA	2820
TAGGTTTTTA	CCTAAAAATC	TTGTGGCAGC	TACATTAGGA	AATGAAGCTG	GTCTGATAGG	2880
AGCTGTAAAA	AATTTCTTAG	ATAGAATTTT	TAATAAATAG	TATGTAAGAT	AAGGAGGTGT	2940
CACAATGACT	AACTCTGTAT	TTTCGACAAT	GCAAGATATT	GAGAATGTTG	CAACCGATAT	3000
TATAAAATCA	TATGATAATG	AGATTTATAC	TTATAAAGCT	GTTTCCCAAG	AAGAATTGGA	3060
AAAACAGAA	AAAAGTTATG	ATGAAAAAAG	TCACGAAGAA	TTAGTTTCAA	TAGAAAGCAA	3120
TTTAGAAATG	AAACAACAGA	ACCTTATTGA	TGAGGTTAAT	AAAACAATCA	AGGAAAATGA	3180
TGCAAAATATT	CAGTATATTT	CATCAAGTAG	GAGAGGAGAA	TTTGTAGAAA	AAATTATTGG	3240
TAGGGTGGTA	GAAAAATATG	GCCATTAGTC	AGATGAAAAG	AATCTCTCTA	CTATTTTCTA	3300
AAAGTAGTCT	TGATGATGTT	TTAAAACTA	TTCAAGAACT	AGAGTCAGTG	CAGTTCCGTG	3360
ATTTAAAGGT	TCAGGATAAC	TGGTCAGAAG	CTCTAGAAAA	AGATGAAGTT	GTATTTCCAA	3420
CTATTCAAAT	TTTTCATACT	TCTAATTCCA	ATCATGGGGT	TATTGAGGGA	AATGATGCCT	3480
TGACTTATTT	GATGAATCAA	CAACAACATT	TAGAAGCAAC	TGTAGAGAAA	TTACAAGAAT	3540
ACCTACCGAA	AGAAAACACG	TTTAAATTAT	TGCAGCAACC	TCCGATAACT	ACCTCTTATG	3600
AAGAATTAGA	GAAATTTGGT	AAAGCTAATG	TTGCTGAGGG	TGTTCTTAAA	AAAGTGAATC	3660
ATCAAATTAA	CAGAGTTCAT	GAATTAGAAA	GACACATTCA	AAGTAATAAT	GAGGAAATAG	3720
AGCGATTAAT	AAAGTGGGAA	AAATTAGAAA	TTGTTCTCTG	GAATTTAGAA	CAATTTTCTT	3780
TCTGTAAAGG	AAAAGTCGGA	ACAATTCCAA	GGACTGAAGA	TAATCGCTTA	TACAATAGTC	3840
TTTGTAGAAA	CAATATTGAA	GTTCAAGAAA	TATTTTCTAA	TGATAGAGAG	TACGGTGTGT	3900
TTGTTTCTA	TCAGTCTAGT	TACTCTATAG	ATTTTGATGA	ATACTTATTT	GAACCATTTG	3960

1254

ATTATTCTAG AAAGGAATTA CCGAAGCAGC GAGTAGTAGA TTTAGATCAA GAAAACATGC	4020
AGTTAATAAC TGAAAAAGAG AATATTATCG CATCGTTGCA AGATTCAAAG AAATATTTGA	4080
TAGATTTACA ATGGCAAATA GACTATATTT TATCTATCTA TGCTCGTCAA ATCTCTAAGA	4140
ATAACTTTTT GTGCACTCCG CATCTAGTTG CATTAGAAGG ATGGATAGAA GAAACTCGTA	4200
TTTTATATTT TATAAAAGTT ATGGATGAGC ATTTTGGACA TTCTATTTAT ATTTATGAAT	4260
CGGAAACATT GACGGATAAT CAAGATGAAA TACCTATCAA ATTAACGAAT CATTCTTTAA	4320
TTGAACCATT TGAATTATTG ACAGAAATGT ATGCTCTGCC CAAATATTAT GAGAAAGATC	4380
CTACACCTGT ATTAGCACCA TTTTACTTTA CATTTTTTGG AATGATGGTT GCTGATTTAG	4440
GCTATGGTTT ACTATTGTTT TTAGGAACAA TGTTAGCATT AAAAATTTTT CATCTACCTT	4500
CAGCAACTAA GAGATTTTTA AAATTCCTTA ATATATTAGG GGTAGCCGTT GCAATTTGGG	4560
GTGGAATCTA TGGCTCATTT TTTGGATATG AGTTGCCATT TCATCTGATA TCTACAACCT	4620
CTGATGTCAT GACTATATTA GTAGTGTCAG TTGTGTTTGG GTTTATTACA GTATTTGCAG	4680
GTTTGTTAGC TTCAGGACTA CAAAAAGTAA GAATGAATAA ATATGCAGAA GCATATAATT	4740
CAGGATTTGC GTGGTGTGTT ATTCTG	4766

## (2) INFORMATION FOR SEQ ID NO: 234:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2484 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

CCTTTTAGAA AAAATTAAAG AATACGACAC CATTATCATT CATCGTCATA TGAAACCAGA	60
CCCTGATGCC TTGGGAAGTC AGGTGGGATT GAAAGCCTTG CTGGAACATC ATTTCCCAGA	120
AAAAACCATC AAAGCCGTCG GTTTTGATGA ACCAACTCTT ACTTGGATGG CTGAGATGGA	180
TCTTGTTGAA GATAGAGCCT ACCAAGGCGC ACTTGTCATC GTCTGTGATA CAGCTAATAC	240
TGCTCGTATC GATGATAAGC GCTATAGTCA AGGTGATTTT CTCATTAAGA TTGACCACCA	300
TCCAAATGAT GATGTATACG GTGACCTGTC TTGGGTCGAT ACTAGTTCAA GTAGCGCTAg	360
aGaTGATTAC CCTATTTGCC CAAACAACCC AACTAGCCTT GGCAGATCGC GATGCTGAGT	420
TGCTCTTTGC AGGAATTGTC GGTGATACAG GTCGCTTCCT CTACCCTTCT ACCACTGCAC	480
GGACTCTTCG CCTGGCTGCT TATTTGAGAG AACATAACTT TGACTTTGCG GCTCTCACTC	540
GCAAAATGGA CACTATGAGC TACAAAATTG CTAAACTGCA AGGCTACATC TACGACCATC	600

1255

TGGAAGTGGA TGAAAATGGT GCTGCTCGCG TTATCCTGAG TCAGAAAATC TTGAAACAAT	660
ACAATATAAC CGATGCTGAA ACTGCGGCCA TTGTAGGTGC ACCTGGACGC ATTGACAGAG	720
TGAGTCTCTG GGGAATTTTT GTCGAACAGG CTGATGGCCA CTACCGAGTT CGCTTACGCA	780
GTAAAGTCCA TCCTATCAAT GAAATTGCCA AGGAGCATGA TGGTGGAGGC CACCCTCTAG	840
CAAGTGGTGC TAATTCCTAT AGCCTAGAAG AAAACGAAAT CATCTACCAA AAGTTAGAAG	900
ACTTGCTTAA AACTGATAA AATACTTGCC AACTTTTTCA GAATCTGATA GACTAGTATA	960
GTAACAATCT ATGGCTCGCA AAGAGACCAT GGCAGAAAGG AAATATTGCA AAATGAAAAA	1020
AGATATCCAT CCAGAATATC GCCCAGTTGT CTTTCATGGAC ACAACTACTG GTTACCAATT	1080
CCTTAGCGGT TCAACAAAAC GCTCTAACGA AACAGTTGAG TTCGAAGGCG AACTTACCC	1140
ATTGATCCGT GTGGAAATTT CATCAGACTC ACACCCATTC TACACTGGAC GTCAAAAGTT	1200
CACTCAAGCA GATGGACGCG TGGATCGTTT CAACAAAAAA TACGGTCTCA AATAATGATA	1260
AGAGAACAGT TTTGGCTGTT CTTTTTTGTT TCTTGAAATC AACTGCTGTT TTCATGTTCC	1320
AGACTCATCT GTAGGTTCTGA TTTCCATGCT ACTAGGCAGG AAGGAAATAG CTGTTTCAAC	1380
ACGTCCATAA TGAGCTATAC TATTGTCACG AACCACACTT TCATTGATGG TCCAAGTGA	1440
ATTCATTTTC TTAAGAGCTT CTCGGACTTT TTCCAAATCT TTGGAGGCAA TGGCCTGCTC	1500
TAAGGTTTCA AAACGAGGAC TTATACTCAT CTGCTTTCAA AAAGCATTCT AGTCCATCTC	1560
CGATTACCGA TGGACTTTAT CACCTCCTTC TCCAGTCCTT GTATGACATC TTGAAGTTGA	1620
TTCATGACAT CTTCCAAAGT TCgAAAGGCT TTATTCTTAA ATCCACGTTT ACGAATCTCT	1680
TTCCACACTT GTTCAATGGG TTCATCTCTG GTGTGTATGG AGGAATAAAG GTAAAATCAA	1740
TATTAGTCGG AATATTTAAG GTACTTGATT TATGCCATAT AGCATTGTCC ATAACGAGTA	1800
AAAGGATAAG CTTGTGAAAG CTCTTCTAAA AAGGCGTTCA TCCACACTCC TTTTATAAAA	1860
CCTGAAATAA GGCATCAATT GTAACAAATT CTCCTGCCTC TGTAGCCTTC AAATGACGGG	1920
CAAGAAAGGC TTTCTCTTCC TCAACTGTCA TATATGCATG GTTACGACCA CCACGTGTTT	1980
CTTGAAGGAG AGAGTCGAGT CCGAACTCCT CATATTTTTT TACGTTTCGC CAAATCGTTG	2040
TTTGATTACA GTCTAAAAGC TCTATAATCT CTTTATAAGA TTTGCCCATC AGACGAAATA	2100
TAGTAGATTG AACTAGAAAT AGTACACCTC TACTTCTAAA ACATTGTAG AAATCGATTT	2160
GTCCTGTTCT TGTTTCATTT TACTATAGAA CGATTTGAAG GCGTTTATAA TATTTAGCTG	2220
TACGAGAGTC TTTTAAAAGT GTTTTGATGG TTTGGATTTT TTCTTTAGTT GATTTTCATAT	2280
TACTATTATA TAATGCTTTT TGATTTTAGT CTGGTATAAA TATTGCTTTC CTCCAAAATG	2340

1256

GTCATAGTTT TACTGGCAAA TCTAACATAT CACGGATAAA TTAACAAGTG ATTTCTGAAT	2400
TGCTAAACAT TTTCTTTTCT TATAGCATAC TTAAAGATTT TGTCTTTGAG AAAGATATTT	2460
CCAAGAAAAA CGTTCGTTTT TTGG	2484

(2) INFORMATION FOR SEQ ID NO: 235:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1766 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

CTAGATATAG CTATAATTTT ATTTATAACA AGAGGATAGA AATGACCGAA TTAGAAAGAA	60
AAAATCGAAA AATTAGCTAA GAAATATTCT GATAACTTAA ACATCAAAGT TCAAGAGAGA	120
GTTCTGTAAA TGGCAAATGA TAATAAGAGC CATTATTTGA TATACAGAGT TTTAGGTATT	180
TCATTTGAAG AAGGAGAAAA TATCGATTTG TATCAAAATA AAGGTCGTTT TTTATACAAA	240
TATGCTGGTT CATTTTTAGA AGAAGCTGCA GTACTATGCT TTAACGAAAA ATTTGGTACA	300
GAAAATACTT AAAAAGTTAA CATTCCTAAT TCTGAAAGTA CAAAACCTAA GACTTTTGAA	360
ATTGATTGTT TACTCGGAGA AAAACACGCA TACGAAATAA AATGGTGGGA TGCAACTACA	420
GATGGAGACC ATATACTAA AGAACACACT AGAATAAAAG TTATTCATAA CAAAGGATAT	480
ATACCAATTC GGTTAATGTT CTAATATCCA AATAGAACTC AAGCTATAAA AATTCAGCAA	540
ACTTTAGAAA CATTGTATAA CGGTATTGGA GGGAAATATT ATTATGGAGA TTCTGCCTGG	600
GAACATTTAA GAGCAGTGAC CGGTATTGAT TTAATTAGTA TTCTAACAGA TATTGCAAAT	660
AAAAAACAG GGGTAAATC AAAATGACAG TATTAAGAGG AGATAACTTA GAAATATTAA	720
AAACTATTGA ATCCTCAAGT ATTGATTAA TCTATATGGA CCCTCCTTTC TTTACACAGA	780
AAACCCAAA ATTATCTAAT AACAAAAATA TTATGTATTC ATTCGAAGAT ACGTGGACT	840
CGATTGAGGA TTACAAAGAA TTTTGTCTG TAAGATTAGA AGAATGCAAA AGAGTGCTAA	900
AAAATAGTGG CAGTATTTTC GTTCATTGTG ATAAAAATTGC AAATCATCAT ATTAGATTAA	960
TTTATAGATA TATCTTTGGA GTAGATATGT TTCAAAGCGA AATTATATGG AACTATAAAC	1020
GGTGGTCTAA TTCAAAAAAG GGATTATTGA ACAATCATCA AAACATTTAC TTTTATTCAA	1080
AGTCAAAAGA TTTTAAATTT AATACAATTT TTACAGAGTA TTCTTCTACT ACAAATATCG	1140
ACCAATACT AGTGAACGA AACGAGATG GAACTCTAA AACTATATAT AAGGTTGATA	1200
ATAATGGTAA CTATATTCTA GCAAAAGAGA AAAATGGAGT TCCCCTTTCA GATGTTTGA	1260



1257

ATATACCATT TCTTAATCCA AAAGCTAAAG AAAGAGTAGG TTATCCTACA CAAAAACCTA	1320
TTCTGTTATT AGAACAAATT ATAAAGATTG CTAAGTATAA AAATGATATA GTTTTAGACC	1380
CGTTCTGTGG AAGTGGAAGT ACTTTAGTAG CCTCCAAGAT TTTGAATAGA AATTATATGG	1440
GGATTGATTT ATCTGAGGAA GCTATCAATA TAACTCAGCA ACGTCTGGAA AATGTTATAA	1500
AAACAAGTTC AAATTTATTG AATAAAGGAA TCGAAGCATA TAGAACCAAA ACTGAGGAAG	1560
AGGAAAACAT TCTTAAATTA TTACAGGCAA AAATGTTTCA AAGAAATAAA GGAATTGATG	1620
GTTTTTTACC TAAACATTTT CAAAAAAAC CGATACCTAT AAAAATTCOA AAAAATAATG	1680
AATGCTGAA TGAGAGTATC TCTTTATTAC AGAATGCTAT AAATCCAAA AAATTTGATT	1740
TTGGAGTAGT TATAAAACT CATTCT	1766

(2) INFORMATION FOR SEQ ID NO: 236:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 748 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

CCGAAATCA AATTCAAACC ACGTCAACGT CGCCTTGCCG TACTCAAGTA CAGCCTGCGG	60
CTAGTTTCCT AGTTTGCTCT TTGATTTTCA TTGAGTATTA AACTAAATTA AATAATATTA	120
GCGCGGAGAA TTTCTAATTC TTCCTTGGTC AAGCGACGCC ATTCCCCTCG TTCTAGGTTT	180
TCATCTAATA CTAAAGTTCC CATAGTCAAT CGTTGCAAGT CCACCACTTC CTTGCCACAG	240
TAGCCACCA TACGCTTGAT CTGATGAAAC TTCCCTTCTG CAATGGTCAC ACGGATTTGG	300
CTTTGATTCT TTTCTGTATC TATGGATACA AGCTCCAGTA TAGCGGGTTG ACAGGTAAG	360
TCTTTGAGAG GAATACCCTC AGCAAATGTC TCCACATCTT CTTGGGTCAT GATTCCCTTG	420
ACTTGTGCCA GATAAGTCTT GTCCACATGA CGCTTGGCG AAAGAAGAAC ATGAGCCAGC	480
TGACCATCAT TGGTCAAGAG CAAAGACCA TGCCTGTCAA TATCCAAGCG TCCTACTGGG	540
AAAATTCCT TACTCCGCGC CAAGTCATCC AACAAGTCCA GAACGGTTCT GTGCTTGGGA	600
TCCTCAGTCG CTGAGATAAC TCCTTTGGGC TTGTTTATCA TGTAGTAGAC AAATCTTCA	660
TACTCCAACA CTGCCCCATC AAAGCGAATC TCATCTATTT TTTCATCAAT CTGCAATTTA	720
GCTGATTTT CTTTTGACC ATTTACAG	748

(2) INFORMATION FOR SEQ ID NO: 237:

1258

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1449 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

AAAAGATTAC ATTGCAACAA TTGAAAATTA TCCAAAGGAA GGCATTACCT TCCGTGATAT	60
TAGTCCTTTG ATGGCTGATG GAAATGCTTA TAGCTACGCT GTTCGTGAAA TCGTTCAGTA	120
TGCTACTGAC AAGAAAGTCG ACATGATCGT GGGACCTGAA GCTCGTGGAT TTATCGTGGG	180
TTGTCCAGTT GCCTTTGAGT TGGGAATTGG TTTTGGCCT GTTCGTAAGC CAGGTAAATT	240
GCCACGCGAA GTTATTTCTG CTGACTATGA AAAAGAGTAC GGTGTCGATA CCTTGACTAT	300
GCACGCGGAT GCCATTAAGC CAGGTCAACG TGTTCTTATT GTAGTGACC TTTTGGCGAC	360
AGGTGGAACT GTTAAGGCAA CTATCGAGAT GATTGAAAAA CTTGGTGGTG TTATGGCAGG	420
TTGTGCCTTC CTTGTTGAAT TGGATGAATT GAACGGCCGT GAAAAAATTG GTGACTACGA	480
CTACAAAGTT CTTATGCATT ATTAATGAAA ACAGTCCCTA GGGCTGTTTT CTCTACACTA	540
GGATATAAAA ATAGACTATA ACTAGTTAGA GAAAACTAT AATTGAAAAC TATATCTTCT	600
TGCAGTATAA TAAAAGGACT AAGTGTTTGA GATTGTGCTT CAAACATATG CAATTATTCC	660
TGAAAGAGTA CAGTTAGGAG AGGGTTATGC CGATTGGAAT TGATAAAAAA TTGCCAGCTG	720
TTGAGATTTT ACGGACAGAG AATATCTTTG TCATGGATGA TCAACGTGCT GCCACCAAG	780
ATATCCGTCC TTTGAAGATT TTAATTTTAA ATCTCATGCC ACAGAAAATG GTCACAGAGA	840
CCCAGTTGTT GCGCCACTTG GCTAATACAC CCCTACAACT GGATATTGAT TTTCTCTATA	900
TGGAGAGCCA CCGTTCTAAA ACAACTCGTT CAGAGCACAT GGAGACCTTC TATAAACTT	960
TTCTGAAGT CAAGGATGAG TATTTTGATG GGATGATCAT CACGGGTGCT CCAGTTGAGC	1020
ATTTACCATT TGAGGAAGTG GACTATTGGG AGGAATTTAG ACAGATGCTT GAGTGGTCTA	1080
AGACTCATGT CTATTCGACC CTTCATATCT GTTGGGGGGC TCAGGCTGGG CTTTATCTGC	1140
GCTATGGTGT AGAAAAATAC CAGATGGACA GTAAGCTATC AGGTATTTAT CCTCAGGACA	1200
CCCTAAAAGA GGGTCACCTT CTATTTAGAG GCTTTGATGA TAGCTATGTA TCCCCTCATT	1260
CACGGCACAC GGAGATTTCT AAGGAAGAGG TCTTAAACAA GACCAATCTC GAGATTTTAT	1320
CAGAAGGACC TCAGGTTGGG GTTTCTATTW TGGCCAGTCG TGATTTACGA GAAATTTATA	1380
GTTTTGGTCA TTTGGAGTAT GACCGTGATA CTTTGGCAAA AGAGTATTTT CGAGATCGTG	1440
ATGCAGGTT	1449

1259

## (2) INFORMATION FOR SEQ ID NO: 238:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 904 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

TACCCGCTTC TTTCAAGAGT TGGAGCAGGG CTTGTTTGCG ATCTTTTGTC ATAGTTCTTC	60
CTTTTAACGG CGTTTTCGAA GCACTTTATA GACAGCTAGT GCTAATGTAT AGTCTACCAT	120
ACTATGGATA ATTGTACCAA ATCCAAC TAG TACAAATAGA ACATAAAACA TATTTTCTAC	180
ATTGGTACCA GAAGTTGCGT AAAAAACGAC ACAGGCCAAT ACTTCAGCAA GGGCATGAAC	240
AACAGCCAAA ACAAAGTTGA AAATCCAGGA AGATTTTGGT TTATCTAGGG TATCGGGGAA	300
TTTTTG TAGG TAAAGAGCTC CTAAAGCACC AAAAGATATA TGGGAAAAAG CCCGAAAAAC	360
GATAACCATG GGATAGCCAG CCATCAAAAA TCCAAACTA GAGGCTAGGA TGACAAAAAC	420
TGCCATCAAG GCGGACAAGA ACATGGCTAT AAAAATAGCG ATGTGGCTCC CCAAAGTATA	480
GGAAGCAGGT GGAATGACAA TCTTGAAAGG CATAACAATT GGAATCAAAA TCGCAATAGC	540
CGTTAAAAGG GCTGTCATTG TCATAAATTG TGTCTTTTTC CGTGTATTCA CAAGAATCTC	600
CTTTTAACT GCATATACAC TAGTATGGTA CAATAAACCA GACAATAAAG CAAGAATTTA	660
CTTGGGTTTA TAGATCATTT TTTAGTTAAA AGTTATAGTA GATTGAAACT AGAATAGTCC	720
ACCTCTACTT CTAAAACATT GTTAGAAATC GATTTGGCTG TCCTGATCGA TTTGTCCTGT	780
TCTTATTTTCG TTTTACTATA GTAAAGATT TATTAAAAAG AACTGTATA GAGCAAAATC	840
TCCACCTTCA GGTTCGGAAA GCGGAGATTG TTTTATTATT TTTCCAGGGT TTGTAGTCGT	900
GGA	904

## (2) INFORMATION FOR SEQ ID NO: 239:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 946 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

CACTCAAAACA TGACTTATAT CAAGACGGAT GGAATTCAAG ACGATGCCAA TCGCTTGAAT	60
--	----

1260

CGTAACATTC AGTTTGCTGT TCGTGAATTT GCAATGGCAA CAATCTTGAA CGGGATGGCC	120
CTTCATGGTG GACCTCGTGT ATACGGTGGA ACTTTCTTCG TCTTCTCTGA CTATGTGAAG	180
GCAGCTGTCC GCTTGTGAGC CTTACAAGGA CTTCCTGTGA CTTATGTCTT TACCCATGAT	240
TCAATCGCAG TTGGGGAAGA TGGTCCGACT CATGAACCAG TTGAGCATTT AGCAGGTCTT	300
CGTGCTATGC CAAATCTAAA TGTTTTCCGT CCAGCAGATG CGCGTGAAAC GCAAGCAGCT	360
TGGTACCTTG CAGTGACAAG TGAGAAAACA CCAACTGCCC TTGTCTTGAC ACGTCAAAAT	420
TTGACTGTTG AAGATGGAAC AGACTTCGAC AAGGTTGCTA AAGGTGCTTA TGTTGTATAT	480
GAAAATGCAG CCGACTTTGA TACCATCTTG ATTGCGACAG GTTCAGAGGT TAATCTTGCT	540
GTCTCAGCTG CTAAAGAATT GGCTAGTCAA GCGGAAAAAA TCCGCGTAGT CAGCATGCCA	600
TCTACAGATG TCTTTGATAA ACAAGATGCA GCTTACAAGG AAGAAATCTT TCCAAATGCA	660
GTCCGCCGTC GTGTTGCAGT CGAAATGGGT GCAAGTCAAA ACTGGTACAA ATATGTTGGT	720
CTCGATGGTG CCGTTCTAGG TATTGATACT TCGGAGCCTC TGCCCCAGCA CCAAAAGTAT	780
TGGCAGAATA TGGCTTTACT GTAGAAAATC TTGTAAAAGT TGTTGAAAC TTGAAATAAT	840
CCTAAAAATC AGGGCGTAAG CTCTGGTTTT TCTTACCAGA AAAGTAAGGT ACAATCTTGT	900
AAAAGTAGCT GAAATTTGAT ATAGTAGTCC TATGTAAAAG ACAAAG	946

## (2) INFORMATION FOR SEQ ID NO: 240:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2764 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

CGGGGCTCCC TAGTTCTTAG GGAGCTATTT TTGTTTTTTC AAGAAGTTAT CTTCTTGAT	60
TTTATACTCA ATGAAAATCA AAGAGCAAGC TAGGAACTA GCCGTAsTG CTCAAAACAC	120
TGTTTTGAGG TTGTAGATAA GACTGACAAA GTCAGGAACA CATATCTACG GCAAGGCGAC	180
GTTGACGCGG TTTGAAGAGA TTTTCGAAGA GTATTAGTTG TGAATCTGGT GCAGTCGTCC	240
CAGATTATTC TTATTAGTAG GGTCTTGTTT TCTATATCCC CTCGTAGTTA ACAAGACCTT	300
GAGCATTTTA GAAAGAGGAA TCTATGTCTA CGAAATATAT TTTTGTAAT GGTGGTGTGG	360
TATCGTCCAT TGGGAAAGGG ATTGTGGCAG CGAGTCTAGG CCGTCTCTTG AAAAAATCGTG	420
GTCTCAAAGT AACCATTCAA AAGTTTGACC CTTATATCAA TATTGATCCG GGAACCATGA	480
GTCCTTACCA GCACGGGGAA GTTTTTGTGA CAGATGACGG AGCTGAGACA GATTTGGACT	540

1261

TGGGTCAC TA TGAACGTTTC ATCGATATCA ATCTCAACAA ATATTCCAAC GTGACAACTG	600
GGAAAATTTA CAGTGAAGTT CTTCTGTAAG AACGCCGTGG AGAATACCTT GGGGCAACTG	660
TTCAAGTCAT TCCTCATATC ACAGATGCTT TGAAAGAAAA AATCAAGCGT GCCGCTCTAA	720
CGACCGACTC TGATGTCATT ATCACAGAGG TTGGTGGAAC AGTAGGAGAT ATCGAGTCCT	780
TGCCATTCTT AGAGGCTCTT CGTCAGATGA AGGCAGATGT GGGTGCGGAT AATGTCATGT	840
ATATCCATAC AACCTTGCTT CCTTACCTCA AGGCTGCTGG TGAAATGAAA ACCAAACCAA	900
CCCAACACTC TGTCAAAGAA TTGCGTGGCT TGGGAATCCA ACCAAATATG TTGGTTATTC	960
GTACAGAAGA GCCAGCTGGT CAAGGAATTA AAAATAAACT GGCCCACTTC TGTGATGTGG	1020
CACCAGAAGC CGTTATCGAA TCGTTGGATG TTGAACACCT TTACCAAATT CCACTGAACT	1080
TGCAGGCACA AGGGATGGAC CAAATTGTTT GTGATCATTT GAAATTAGAC GCACCAGCAG	1140
CGGATATGAC AGAATGGTCA GCCATGGTGG ACAAGGTCAT GAACCTCAAG AAACAAGTTA	1200
AGATTTCCTT TGTGGTAAG TATGTGGAGT TGCAAGATGC CTATATCTCA GTGGTCGAAG	1260
CCTTGAAACA CTCTGGCTAT GTCAATGATG CAGAAGTTAA AATCAATTGG GTCAATGCCA	1320
ATGATGTGAC AGCAGAGAAT GTAGCAGAAC TCTTGTCTGA TCGGACGGG ATCATCGTAC	1380
CAGGTGGTTT TGGTCAACGT GGTACAGAAG GGAAATCCA AGCCATCCGC TATGCGCGTG	1440
AAAATGATGT TCCAATGTTG GGAGTCTGCT TGGGAATGCA GTTGACATGT ATCGAGTTTG	1500
CTCGTCACGT TTTAGGTCTT GAAGGTGCCA ATTCTGCAGA GCTTGCACCA GAAACAAAAT	1560
ACCCTATCAT TGATATCATG CGTGATCAGA TTGATATTGA GGATATGGGT GGAACCTTC	1620
GTTTGGGACT TTATCCGTCT AAGTTGAAAC GTGGCTCTAA GGCTGCTGCT GCTTATCACA	1680
ATCAAGAACT GGTGCAACGC CGTCACCGTC ACCGTTATGA GTTTAATAAT GCCTTCCGTG	1740
AGCAGTTTGA GGCAGCAGGT TTTGTCTTTT CAGGAGTTTC TCCAGACAAT CGTTTGGTAG	1800
AAATCGTGA AATTCCTGAA AATAAATTCT TTGTAGCTTG TCAGTATCAC CCTGAACTGT	1860
CAAGCCGTCC AAACCGACCA GAAGAACTCT ACACTGCCTT TGTTACTGCA CGAGTTGAGA	1920
ACAGCAATTA GCAAAATCAG AACCTTTGAG AAAAATCTCA GAGGTTTTTT GCATACGATG	1980
ATATTGCAGT ATATCTGAGG TAGGGTCCT CTGTATGTAC CTGCTACCGT TGAAATCAAT	2040
AGCGACTCCC TCTTGCCCTG TGCTAGTGAA TGGATTTATC AGTATATTGA AATGAAATAA	2100
AATTTGAACA AATTAATTCT GAAAGCCAAA TCAATTTCTA GCAAAGTTTT AGGAACTGGA	2160
TTGTATAGTG AATTGAAATA AGATGTGAAC ATCTCTATCA GGAAAGTCAA ATTAATTTAT	2220
AGAAATATTT TAGCAGTCAA GATGTAATGT TATAGATTCA ATACATTATA CTTTTTTAAT	2280

1262

TTAATCCACT ATAGTAAAT GAAATAATAA CAGGACAAAT CGATCAGGAC AGTCAAATCG	2340
ATTCTAACA ATGTTTGTAGA AATAGAGGTG TACTATTCTA GTTTCATAT ACTATCCCAA	2400
ATCATTGATA CCTCTCTCAA CTAGATGTAA CTTACAAAAC CCCTGACCTC ATGAGCCACT	2460
TTCTTCCTCC TCATGAGGTC AGTTTACTT TCTGCTGTC CAGTATCGTT TTTCCTCGCT	2520
AGATTTCCTC AAAAGGGCAG ACTCCTCCCT TGGTGGCTCA CACGATTTT TCATCTCGAC	2580
TGTTCTTTAA TGCATCATT ACGACGCTT TCTTCTAGGT GGTTCATAAG GAACAGGAAG	2640
ATTCAGGTG ACTTTTCTAA TCCTAGAATA AAGTGTGAA AACAAATCGG AATAGGCATA	2700
GAGACTAGAC AATTTGAGGA GCTGCTGCG TCCTGTTGGA ACACATTTT CCACCACGTG	2760
AAGA	2764

(2) INFORMATION FOR SEQ ID NO: 241:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1682 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

CCGTTTTTTT CATGTTTCAG TACTACAAC TACGTTGTAG CGCCCTGCAC ATTGGTTCGT	60
CTTGTTTCAGT TTTCAAAGGT CTTGTGCACT TGCTTCTCTC AAGCGACAAC TATATTAGTA	120
TATCACAAC GCTTTCGCTT GTCAACACTT TTTGAAGAT TTTTAAGTTT TTTTAAACTT	180
TTTTTCATCA AGTGGTCCTG ACGCAACATA CCATAGTCCG TACGGGATTC GAACCCGTGT	240
TACCGCCGTG AAAAGGCGGT GTCTTAACCC CTTGACCAAC GGACCTGAGT TGTATTTTTC	300
AACTCTTACT ATTATACAGT CTTTCAAAC TTTGTCAACT ACTTTTTTAA ACTTTTTTTA	360
TTAATTTTAC AACAGCTTCA GTTCGAGCTG TATGTGGGAA CATATCGACC GACTGGATAT	420
AATGAAGATC ATAGACTTCT ACTAAGCGTA CCAAATCACG AGCCAAGGTC GAAACATTAC	480
AAGAAATATA AACCATTTTT TCTGGTACAT AAGTAAGAAT AGTATCTAAT AACTTATCAT	540
CCAGACCTGT ACGTGGTGGG TCAACAATCA AAGCATCTGC TCGGTAGCCT TCCTTGTAAC	600
AACGAGGAAT AATCTCTTCT GCCGTCCAG CTTCATAATG AGTATTGTCA AATCCCATTC	660
TTTTAGCATT TCGCTTGGCA TCTTCAATAG CTTCTGGAAT AATATCCATA CCTCTGAGTG	720
TTTTTACTTT CTTTGCAAAG GCAAATCCAA TCCTTCCAAC TCCACAATAA GCGTCAATCA	780
AATGGTCTTC TTTATCAACA TCCAGCGCTT TTAAGCTTTC GCTATAGAGG ACTTCTGTTT	840
GCTCAGGATT TAGTTGATAA AAAGCTCGAG GGGATAGTGA AAATTCATAA TTGAGTACAC	900

1263

CTTCTTGAAT ACTCTCTTGC CCCAGATAA TCTCTGTCTT TTCACCATAT ATCTCACTGG	960
TTTTAGCTGT ATTTGTATTA ACAGCTACTG TCACAACTTC TGGGAAATCT TTAACCAACT	1020
CTTTTACCAA TTGAGTTAAA TTAAGCTGGC GGTTCGTAAC AATAATAATC TGAACCTGTC	1080
CGGTCTTTCT CGCGCGTCGG ACCATAATAG TACGGACACC TAGAACTTTT CTCTCATCCG	1140
TGATTGGAAT CTGGTGATAA GTAAGTAATT CTGCTAAGCG ATTAGCAATC ACTTGGGTTC	1200
CCTTATCTTG TACCAGGCAG TCTTTCAACT CTAATAATA GTGAGAGTTT TGTGCATATA	1260
AGCCCCGCTT GACCTGATTT TTAATTTTC GAGTCTGAAA TTGTAACCTA GCTCTGTAAT	1320
ATTTTGGTTC CTGCATTCCA ATAGTTGGAC GAATTCATA ATTTTCATAT CCTGCAGGAG	1380
CAAATTTTT CAGCGCTTGA TGAAGTAAGT CCGTCTTGAA CTCCAGCTGC TTATCATAAT	1440
GCAGGTGCAT GATTTGGCAG CCTCCGCATT CATTATAAAT AGTACAAGAT GGCACAATTC	1500
GAAATTTAGA CTTCTTGTG ACCTTCAGTA ATTTTGCTTC AACAAAGTTG CGTCTAATAG	1560
AAGTAATCTG ACAATAGATA TCTTCGCCTT TGAGAGCTCC TGGTACAAAG ACTAATGTTT	1620
TTTGGTAAAA GCCGATTCCC TCACCGTTAA TTCCCATGCG CTGATTTTT AATGGTATTT	1680
TT	1682

(2) INFORMATION FOR SEQ ID NO: 242:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2524 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

TTAACTTTGG TCAATTCMTT AAAGTCATCC TCTGTAAGCA TGTCTAACCA TTGATGTTTC	60
CCTTTATTGC TAAATCACC AATTCGACT ACAGCTATAT CTAAATCTTT CCAACTATTT	120
TTCAAATTTT CAAATATCT TGATTGCAAA ATACCATCTG CTAACAATTT ATTTCTTGC	180
ACAATCGTTG CATTCATAAA TGTACACTCT CCATGAAAT TTCTAGACAT TTCATAAATC	240
AGTGTATTCA CATGGTATTT AGCGTGTATG TGAAGTAGGAC CACCTGCTAG AGGATAGAAG	300
TGAACATTTT GGACACTTTT ACTGTGAATT AAATCTACTA AATTACTTAA ACTTTTCCCC	360
CAAGAAAAGC CAATTTTCAT ATTATCATCA ATTAGATTCC TAAGGACGCC TGCTGCAACT	420
TGAGAAATTC TTTAGATAA AATTGTTGGA GTATCATCAA ATTCATTTGG AATAATTTCT	480
AAACTTTCCA AACTGTATTT TTCTTTTACA TAATTTTCCA ACTTAAACAT ATTGGTATCA	540

1264

AAATTCTCTA TTTC AATTTT AACAAATTCCT ACATTCCTTG CTTCTGTAA CATTCTACTA	600
ATAGAGGTTT TATAAATTC TAATTTTGCT GCTATTTGTG ACTGATTTAA GTTTTCAATA	660
TAATACAGAT AAGCAATTTT AGAAAGCAGT TTATTCCTAT CTTGATTCAT AACTTTAACC	720
TCTTACGAAA CTACCTTAAC CATTATCCCA GCATTTTCTA ATGTAGCTAT ATTTTGTMTA	780
GAAAGTTTTT CGTCTGTAT TACTTCATAG ACTTGACTTA AAGCAAATCT TCTTACTGTA	840
CCTCTTTTAT CAAATTTACT TGAGTCAGTT AGGACAATGA CTTTATCCGA CACTGCTGAA	900
ATATATTGAA CTACCTCACT GCGCATTAAT TCTTTTCCGG TAAAGCCCAT CTCTTTATCG	960
TAACCATCTG TCCCAACAAA AGCTTGACAC ACATGAAAAG TCTGTATCAT TTCTTTTAAT	1020
AAAGGTCCTA CAGTCACCTG TGAATCTTTC TGAAACTCAC CACCAAGAAC AATAACACGA	1080
CATGAATCAT AAGCTCTCAC AAAATTTGCT ATAAAAACG AATTTGTAC AATCGTAACA	1140
TTTCTTTTTT GCTTGCAAAT TTCCTCAGCA AGTAAAGCAC AGGTCGATCC AGATTCTATC	1200
ATTATTGTTT CATTATCTGA CACCAATTTT ACTGCTTCCT GAACAATTTT TCTCTTAGTT	1260
TCATAATTAA TTGACAAACG TACATTTAAG TCATCTCCAC TATTTAATAC AGCATATCCA	1320
TGCTCTCTGT GTAATAAACC TTTTGACTCT AATTTATCTA AATCTTTTCT AATCGTTACT	1380
TTGATACAT TTAATTTTTT CGATAATGTA TTAACGTCGA TCTTTTCATA TTCTGATACT	1440
AATTTAATAA TTTGTTCCAA TCTTTTCATT TTACACCTCC GTTTTATTCT ACCAAAATAA	1500
AAAGCAAAAA ACAACAAATT AACCTTTCGT TCGTAATTGT TTTTCTTTCC TTTTGTGAT	1560
AGGATAGACT TATGAAGAGG AGGAACTCTT ATGGAAATAT CTAAAGGAAT TATTTTAAAT	1620
ATTCAACACT TTTCAATTCA TGACGGTCCG GGTATTTCGT CAACTGTTTT TTTAAAAGGA	1680
TGTCCTCTGC GCTGTCCATG GTGTTCTAAT CCTGAATCTC AAAGAATGAA ACCTGAAAAA	1740
ATGAAAGATG CTCAACGAGA GAAATTCACC TTAGTCGGTG AAGAAAAGAC TGTAGAAGAA	1800
ATTATTACAG AGGTATTAAA AGACAAAGAA TTTTACGAAG AATCCGGTGG AGGTTTAACT	1860
TTATCAGGAG GTGAAATATT TGCTCAGTTT GAATTTGCTA AAGCCATCTT AAAATCAGCT	1920
AAAGAACATC ACATACACAC TGCCATTGAA ACTACTGCCT TTGTTGATCA TGAAAAATTT	1980
ATTGATTTAA TTCAATATGT GGATTTTATC TACACAGACC TAAAACATTA TAATTCTATA	2040
AAACATAAAA AAGTGACTGG GGTTTTAAAT CAAATGATTA TTA AAAACAT TCATTATGCT	2100
TTTTCACAAA ATAAAACTAT CGTTTAAAGA ATCCCAGTTA TTCCTAATTT TAACAATAGT	2160
TTAGAGGATG CAGAAAAATT CGTACTCTA TTAACTCAT TAAATATCGA CCAAGTTCAA	2220
CTACTCCCTT TTCATCAATT TGGTGAAAAC AAATATCGTT TATTAATCG GAAATATGAA	2280
ATGGATGGAA TCAACGCACT TCATCCWGA GATCTTATTG ATTATCAAAA GGTATTTCTG	2340



1265

AACCACCATA TTAATTGTTA TTTCTAGTTT ATTCCTTGA AATGCTCTAG CTATTTGCAG	2400
ATAACAAGCA TCTATAATAC ATACTTAAC TTTCAAAAGG TTTAGCTAAA AAATTTTAGC	2460
CAAACCTTTT CTATTTTACC TTGCTCTAGA ATTTTAAAC TGCTATACTT ATCACAAAAA	2520
AACG	2524

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2359 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

CGTGCTTGGG GGCTTGTGGT CAAAAGGAAA GTCAGACAGG AAAGGGGATG AAAATTGTGA	60
CCAGTTTTTA TCCTATCTAC GCTATGGTTA AGGAAGTATC TGGTGACTTG AATGATGTTT	120
GGATGATTCA GTCAAGTAGT GGTATTTACT CTTTGAACC TTCGGCAAAT GATATCGCAG	180
CCATCTATGA TGCAGATGTC TTTGTTTACC ATTCTCATAC ACTCGAATCT TGGGCAGGAA	240
GTCTGGATCC AAATCTAAAA AAATCCAAAG TGAAGGTCTT AGAGGCTTCT GAGGGAATGA	300
CCTTGAACG TGTCCCTGGA CTAGAGGATG TGGAAGCAGG GGATGGAGTT GATGAAAAA	360
CGCTCTATGA CCCTCACACA TGGCTAGATC CTGAAAAAGC TGGAGAAGAA GCCCAAATTA	420
TCGTGATAA ACTTTCAGAG GTGGATAGTG AGCATAAAGA GACTTATCAA AAAATGCGC	480
AAGCCTTTAT CAAAAAGCT CAGGAATTGA CTAAGAAATT CCAACCAAAA TTTGAAAAAG	540
CGACTCAGAA AACATTTGTA ACACAACATA CAGCCTTTTC TTATCTAGCG AAGAGATTTG	600
GGCTTAATCA ACTTGGTATT GCAGGTATCT CTCCTGAACA AGAACCAAGT CCACGACAAC	660
TAACAGAAAT TCAGGAATTT GTTAAGACCT ATAAGGTAA AACGATTTT ACAGAAAGTA	720
ACGCTTCTTC AAAAGTAGCT GAACTCTTG TCAAATCAAC AGGTGTGGGT CTTAAACTC	780
TGAATCCTTT AGAGTCAGAC CCACAAAATG ACAAGACCTA TTTAGAAAAT CTTGAAGAAA	840
ATATGAGTAT TCTAGCAGAA GAATTAAAGT GAGGAAAGAA TGAAAATTAA TAAAAAATAT	900
CTAGCAGGTT CAGTGGCAGT CCTTGCCCTA AGTGTGTTT CCTATGAGCT TGGACGTTAC	960
CAAGCTGGTC AGGATAAGAA AGAGTCTAAT CGAGTTGCTT ATATAGATGG TGATCAGGCT	1020
GGTCAAAAGG CAGAAAACCT GACACCAGAT GAAGTCAGTA AGAGGGAGGG GATCAACGCC	1080
GAACAAATTG TTATCAAGAT TACGGATCAA GGTATGTGA CCTCTCATGG AGACCATTAT	1140

1266

CATTACTATA ATGGCAAGGT TCCTTATGAT GCCATCATCA GTGAAGAGCT CCTCATGAAA	1200
GATCCGAATT ATCAGTTGAA GGATTCAGAC ATTGTCAATG AAATCAAGGG TGGTTATGTC	1260
ATTAAAGGTAA ACGGTAAATA CTATGTTTAC CTTAAGGATG CAGCTCATGC GGATAATATT	1320
CGGACAAAAG AAGAGATTAA ACGTCAGAAG CAGGAACGCA GTCATAATCA TAACTCAAGA	1380
GCAGATAATG CTGTTGCTGC AGCCAGAGCC CAAGGACGTT ATACAACGGA TGATGGGTAT	1440
ATCTTCAATG CATCTGATAT CATTGAGGAC ACGGGTGATG CTTATATCGT TCCTCACGGC	1500
GACCATTACC ATTACATTCC TAAGAATGAG TTATCAGCTA GCGAGTTAGC TGCTGCAGAA	1560
GCCTATTGGA ATGGGAAGCA GGGATCTCGT CCTTCTTCAA GTTCTAGTTA TAATGCAAA	1620
CCAGCTCAAC CAAGATTGTC AGAGAACCAC AATCTGACTG TCACTCCAAC TTATCATCAA	1680
AATCAAGGGG AAAACATTTT AAGCCTTTTA CGTGAATTGT ATGCTAAACC CTTATCAGAA	1740
CGCCATGTGG AATCTGATGG CCTTATTTTC GACCCAGCGC AAATCACAAG TCGAACCGCC	1800
AGAGGTGTAG CTGTCCCTCA TGGTAACCAT TACCACTTTA TCCCTTATGA ACAAATGTCT	1860
GAATTGGAAA AACGAATTGC TCGTATTATT CCCCTTCGTT ATCGTTCAAA CCATTGGGTA	1920
CCAGATTCAA GACCAGAAGA ACCAAGTCCA CAACCGACTC CAGAACCTAG TCCAAGTCCG	1980
CAACCAGCTC CAAGCAATCC AATTGATGAG AAATTGGTCA AAGAAGCTGT TCGAAAAGTA	2040
GGCGATGGTT ATGTCTTTGA GGAGAATGGA GTTCTTCGTT ATATCCCAGC CAAGGATCTT	2100
TCAGCAGAAA CAGCAGCAGG CATTGATAGC AAAGTGGCCA AGCAGGAAAG TTTATCTCAT	2160
AAGCTAGGAA CTAAGAAAAC TGACCTCCCA TCTAGTGATC GAGAATTTTA CAATAAGGCT	2220
TATGACTTAC TAGCAAGAAT TCACCAAGAT TTAAGTGATA ATAAAGGTCG ACAAGTTGAT	2280
TTTGAGGCTT TGGATAACCT GTTGAACGA CTCAAGGATG TCTCAAGTGA TAAAGTCAAG	2340
TTAGTGAAG ATATTCTTG	2359

## (2) INFORMATION FOR SEQ ID NO: 244:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1052 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

TTCTTTCTGC TATAATCGTA TAAATACTT ACTTTAGGAG TTCTTATGAA AGTTGTAAAA	60
TTTGGAGGTA GTTCTCTTGC CTCTGCTAGT CAATTAGAAA AAGTTTAAAA CATCGTCAAA	120
AGCGATTGAG AGCGTCGTTT TGTAGTCGTT TCTGCGCCTG GTAAACGCAA TGCTGAAGAT	180

1267

ACTAAGGTTA CGGATGCCCT GATTAAATAC TACCGCGACT ATGTTGCGGG TAACGATATT	240
AGCAAGAACC AAAGCTGGAT TATCGACCGC TATGCTGCTA TGGTTAGTGA ATTGGGACTA	300
AAACCAGCTG TGCTAGAAAA AATTTCTAAA AGCATTACAG CCTTGGCCAC TCTTCTATT	360
GAAGAAAATG AATTTCTCTA CGATACTTTC CTAGCAGCCG GTGAAAATAA CAATGCCAAA	420
TTGATTGCTG CCTACTTTAA CCAAATGGT ATCGATGCAC GCTATATGCA CCCTAGAGAA	480
GCTGGGATTG TGGTCACAAG TGAACCTGGT CACGCTCGCA TCATTCCATC AAGTTATGAC	540
AAGATTGAAG AATTGACAAA CACCAATGAA GTCCTTGTC TTCCTGGTTT CTTTGGTGTC	600
ACTAAGGAAA ATCAAATCTG TACTTTCTCA CGTGGAGGTT CTGATATTAC AGGTTCTATC	660
ATTGCTGCTG GTGTCAAAGC TGACCTCTAT GAAAACTTTA CGGACGTTGA TGGTATCTTT	720
GCAGCCCACC CTGGTATTAT CCACCAACCA CACTCGATTC CTGAGTTGAC CTACCGTGAA	780
ATGCGCGAGT TGGCCTATGC AGGCTTCTCA GTCCTTCATG ACGAGGCTCT TCTTCTGCC	840
TACCGTGGA AAATTCCTCT GGTATCAAG AATACCAACA ACCCTGACCA TCCAGGTACT	900
CGTATCGTTC TAAACACAG TAATGATGAA TTTCCAGTTG TGGGAATTGC TGGTGACTCA	960
GGCTTTGTCA GCATTAACAT GTCGAAATAC CTCATGAACC GTGAGGTTGG ATTTGGCCGC	1020
AAGGTTCTGC AAATCCTGGA AGAACTTAAC AT	1052

## (2) INFORMATION FOR SEQ ID NO: 245:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 855 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

CCCTCGAAAA CTAAGCCGAT GAAGTCAGAA CACTTCAATC CTGTTCTGTA CTGGTGGGAA	60
AATCGTGAAG AGATTCTGGA AGGTAAGTTC TACAAATCTA AATCATTTAC ACCTAGTGAA	120
TTGGCTGAGT TGAATTATAA TTTAGACCAG TGTGACTTTC CAAAAGAGGA AGAGGAAATC	180
TTAAATCCCT TTGAGTTGAT TCAGAATTAT CAAGCGGAAA GAGCAACTTT AAATCATAAG	240
ATTGATAATG TATTAGCTGA TATTTTGCAG TTGTTGGAGG ACAAATAATG ACACCAGAAC	300
AACTTAAAGC AAGTATTCTC CAAAGAGCGA TGGAAAGGAA ATTAGTGCCG CAAAATCCCA	360
ATGACGAACC TGCAAGTGAA TTATTAAAGA GAATTAAAGC TGAAAAAGAA AAACCTTATCA	420
GTGAAGGAAA AATCAAACGA GATAAAAAGG AAACCTGAGAT ATTTCTGTTG GATGATGGGA	480

1268

AACATTATGG GAAGTTTGCT GATGGAAGCA CTCAAGAAAT TGATGTTCTT TATGATATTC	540
CTGATACTTG GGAGTGGGTG AGGATAAAAT CAATTATTG GAATTTTGGG CAAAATAAGC	600
CAGAGAAATC CTTTAGGTAT ATAGATACGT CTAGTATTGA TAGAAAAAAG AACATAATCA	660
ACTACAAAAA TCTACAATAT CTTTCACCTG AACAAGCGCC TTCCCGTGCT AGAAAATTAG	720
TTTCGCAGAA TAGTGTCTTA TTTTCAACAG TTAGACCATA TCTAAAAAAT ATTGCTGTAG	780
TTAGAGAACT TAAAGAGTAT TTGATAGCTA GTACAGCATT TAATGTTTTG GGATACTTTA	840
CTTAACGAAA CATAT	855

## (2) INFORMATION FOR SEQ ID NO: 246:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 660 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

TTTAGGAAGG CTATCCGTAA TTTTACAAAG GATTTAGATA TTACAGAGGA ACATTTAGAT	60
ATTATCAAAA GAGAGATGTT TGGCGAATTT TTCAGTAGCA TGAAGTCTCT TGAATTTATT	120
GCAACGCAAT ATGATGCTTT TGAAATGGT GAGATAATTT TTGATTTGCC GAAAATTTTA	180
CAGGAAATTA CTTTAGAGGA TGTCTTGAT GCTGGACATC ATTTAATAGA TGATGGTGAC	240
ATAGTTGATT TTACAATATT CCCATCGTAG TAACCTATTA TAATAGACAC TAGAAAGAAG	300
GGATGACAAG TATGAGAAAA AAAACAATTG GAGAGGTTTT ACGATTAGCT AGAATCAATC	360
AGGGATTGAG TTTAGATGAA TTGCAGAAAA AGACAGAAAT CCAGTTAGAT ATGTTGGAAG	420
CAATGGAAGC AGACGATTTC GATCAACTTC CAAGTCCTTT TTACACGCGT TCTTTCTTGA	480
AAAAATATGC ATGGGCTGTT GAGTTAGATG ACCAAATTGT TTTGGATGCT TATGATTCTG	540
GGAGTATGAT TACTTATGAG GAAGTAGATG TTGATGAAGA TGAGTTGACA GGTCTAGAT	600
GTTCAAGTAA GAAAAGAAG AAAAAACAT CATTTTACC TTTATTTTAT TTTATCCTGG	660

## (2) INFORMATION FOR SEQ ID NO: 247:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1805 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

1269

CCGGTTGCAC AGGATCGTGC ATAGTCAACT CTTCAAGTAT AGCATATCTC CTATTTTCTT	60
ACAAGTAATA ACACCTAAAA TGAAGCTTTT TCTTTTACTT TTTTCTGCCA AGAGGCCAAAA	120
AGCATGCTGA GGTAAAAAAC GCTCATCATA ATAGGAACAC CAAGAATGGT CTTTTCATGA	180
TAGAAAATCG TCAAATAGGC TGAAAAGACA ACGCCAAGGA CAAAACACT AAGCAGGCTA	240
ACAAATATGA ATCCTTCACG CAAAAAAGGA GTGTGCTTGG TTCGGAAATA ATCTCCAAAA	300
GCCAGCATGG TCCGTTTGAT ATTCCCTGTC ATAAAAGCGT TATTATAGGC AATACCCGAC	360
ACTTCTCCAA AAGCAGTTGT CACCAGTCCC ATACAGAAGG CCAAGGGCGG CACTAGATAG	420
ATATTATCCA CAGTTTGCGG CACAAAAGCA ATAATGATTG ATAAGATTGC CAAGGGAATC	480
AAGGACAGAA TAGGTTTTTT CACAATTCTC AATTTTTCCT TATAAATCGT TAATAAAAAG	540
ACTCCCATCA TAAACGCTAG CAAGGTGAGA ACCTTGTCCT TAACATCCGA AACATTATTT	600
TTAATTAATT CTAAGTAAAG AAAGACAACA TTTCCAGTTT GTCCAGCTAC AAGGGTATTC	660
CCGCGAACAA TAAAAGTGTA AGCATCCACA TATCCAGCAC AAAACGTCAA AAAAAGTGCT	720
AACCTTTTAG ACTGACGTGA TATTTTCTT ATAGGTAATA ACCTCATTTT ACCTCCCAT	780
GTATTTCTC TTAGAAATAT TGTACCATTT TCTTTCTAAA AAATCGTAGG CTACCATTTA	840
GATTTTACTA TTAGCATAAA AATAATAATA GACAACTATT TATCCAAAA TAGATAGATG	900
TAACATGTTT GCAAACAAAG CATACGAACC TTTAGTAAAA TCATTTCCAT GAAACTAGAA	960
TAGAGCCCTC TTAGCAAAAA TCATTATTTT AATTATTTC TAATCACTCC TTGACATAAA	1020
TAATCTCAC CAATAAAGA CTATGTCTTA AAAAAATGGT ATAATAAAAT CAATACTTGG	1080
GCTTGATGGC TATGCTACTA ATAACAATTA GGAGAGAAAA TCAGGCACTT GTTAACAACA	1140
AGGATTATCC CCTTGAGATG AAAGGAACTT TAGAAATCTT ATGATGAACA TGCAAAACAT	1200
GATGCGTCAA GCACAAAAAC TTCAAAAACA AATGGAACAA AGCCAAGCTG AACTTGCTGC	1260
TATGCAATTT GTTGGCAAAT CTGCTCAAGA TCTTGTCCTAA GCGACCTTAA CTGGCGATAA	1320
GAAAGTTGTC AGCATTGATT TCAATCCAGC TGTGCTTGAC CCAGAGGACC TTGAGACTCT	1380
TTCTGATATG ACCGTTCAAG CCATCAACTC TGCTCTTGAA CAAATCGATG AAACCTACCA	1440
GAAAAAATG GGTGCTTTTC CTGGGAAATT ACCTTTCTAA AAACAAGGAG CTAGAACAAT	1500
GCTTGTCGAT AACAAAGGCT AAGAAAGGTG CAAAAATGAC TCTATAATAT TTGTAGTGGG	1560
TAAATCCCCT ATGGATATTA TGGAGCCTAT TTTTGTGTAG AAAAAAGTCC CATATGACCT	1620
ATAATGAAAA GCGACAAAAC AACTCATTAG AAAGAATCAT ATGGAACAAT TACATTTTAT	1680
CACAAAATTA CTAGACATTA AAGACCCTAA TATCCAGATT TTAGACATCG TCAATAAGGA	1740

1270

TACACACAAG GWAATCATCG CCAAACCTGGT CTATGAAGCT CCATCTTGTC CTGAGTGCGG 1800  
AAGTC 1805

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2516 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

CTGCATCTAG TTTGTTTCTC CCTACAGTTT TAGCTAGACA GATTGGAGAT TATGATTAA	60
CGTCGCCGCG TTGGGGTTCG GATACAATA GTGAGCTTGA GAAAGAAAAC TCCTCTGCTG	120
GAATTAATAA TAATGACAGC ACTGGTGGCG GTAAAAGGTT AAATACCTCT ATTCGTAGCG	180
CCTATAGTGG GTCAGATATT ACCCCGGTAT ATTCATTGGG GTCTGGCTCT AGGATTGTCA	240
TGTACTATAA TGGAGGTGGT GACAATTATA TTGGTTCTGG TACTAGATTA GCTATGGCGC	300
CACAATTTGG AAATCATGTA AGAATTCATA CTTCAGGTTT TTGGAATCCA GATTCTTATT	360
AACTTACTTG TCAGAGTAAG CCTTAAAGAT GGTGATTGT GGGTGTAGCA TGAAAAAGA	420
ATGCTACACC CTATTTTAT TATAAGGAGG AGTAAGGATG GAATTTTCA TTTGTAATCT	480
TGTACGAGTC GTTCAATCAC CTCGATTTTA TATGTCTTTA TTTTGGACCC TTCTTTGCAT	540
GAGTTTAGGA AATTTCCTTG CTTTCAATGG TATTTATAAA ATTGAAGGTT TATCGATTTT	600
TTTTGCCGCT TCTTCTATTC GAGGATTTTC ACCGATTAGC CTAGTAGCTG CACTTATCTG	660
TACACTGCCC TATTCTAGTC AGATAATAGA GGATGCTGAG AGTCATTTTC TAACAGCACA	720
ATTGTGTCGA ATTTCTAAAA AGAAGTATCT GGCTATTGTG GGTAGTACTG TAATTATTTT	780
TTCTTTTCTA GTCTTTTTC TCCCCTATTT ATTATTATTA GGAATTAATC TTTTAGTGAC	840
TCCTTATCAG GAAATTTATA TTGGAGATTA TAGTGGTGCC TTAAGAAGAT TATTTGATTC	900
CAATCAGTTT CTCTATAGTC TTGTAACGAC TCTCTGGTAT GGAGTTTGGG GCGCTGTGTT	960
CTCTATTTT GGACTAGCTA GTGCTTTGCT AGTGAAGAAA AAAATAGGAG CTATTTTCAT	1020
CCCAGTTGCC TATATGATGG TTGGTGGTAT TTTTGGGCT ATTTTAGGGC TATCTTACTT	1080
AGAACCTGTG ACAACGCTAG CTTTGGGATA TCAGAAAGAT ATCAGTCTTT CCTTAGTTAG	1140
TGCTCATCTT GCTTTTATTT TATTTGTTAG TTGTTTGGTT GTTTATGGTA CATTTTCTT	1200
ACATTCAGAG GACTATGTAT AATGAAACAA TTTGTTCAAT TTTATAAAAA AGATTTCTTA	1260
GCAGTATTGG TTTATTTTAT ATTACTGCTA TCCTGTGTTT TATCTAGTAC AGTATATTTA	1320

1271

TTGCGCtGTC GCCAATATTC AATCCATCCA AATGTATTAG AATGGATCTT AGTTTTACTT	1380
CAAGATATGA CGACTGGAGT ATATTGCTTT CCGTTCACAT ATATATTGTT CTTTTTTTAT	1440
TTGATGAATA ACTATTTTAA TAGGTTGGAG TGTCGCATTC GTCTGAAATC AATTAAGCAC	1500
TTTACCAGTT TTAGTTTCAA ATTAGCAGCT CTTAGTACGG GGATTGGAC GCGGACTTTA	1560
TTTTTATTGA TTTTCTAAT TGCATTTAGT AATGGTTTTA GCTTCTCTTT GGAGATAAAG	1620
GAGGTTGATT TTTTAAGAGA ATTTTATGGT ATAAGTATTG CAAACAATGC TACTTTCTTT	1680
ATAGGATTTT TTTTCTCTTA TATAGCATAC TATTTCTTTT TATCCTTACT TACTATTAGC	1740
AGTTTTTCTT GGTTTAAAAA ATCAAACATG AGCTTAGTAT TTCTGTTTAC TTTTTTATTT	1800
GTAGAATCCT TATTCTGGAT TTATCAGTTG GACAATGGGA TAATTGGATT ATTGCCAATT	1860
TTTCAGTATA TGGTAAATTC CAATCCGTAT GCATTGATTT ATTGGCTTAC ATTACTATCT	1920
ATCATAATTC CATTGACTGT ATTTTCTGTT CATAGAAACT GGAGGAGAGT GTAAAAGTTG	1980
GAAATGGGAA AGTTAAGTAG TCACATGTGG AGGTGAATC AGATAATCTA TACCAAGTAC	2040
TTTTGGGGTT ATGTTCTTTT TTGGATATTG ATTTGTTTAG GATTATGGTA TTGGTTAGAA	2100
GGAAATGATA GACTTGTTAT AGAAATTTTA AAAGGGCCTA ATCTGAGTCA AAACCTTTTT	2160
TTAGTCTTAT CTATATGGTT GCTTCATTGG TTTATTATTC ATACATTTTT TCTAGCAGTT	2220
GTATATCGTA GAAGAGCATC CGATTCTTTT ATGGAAGTGA TTCGATTTTC TTCTATTAAG	2280
CTCTGGATTA GGTATCAGAT TTGGACCTGT TTTCTTTATG GACTCATTTT AATCATGGTA	2340
AAAGTTCTAG TGATTCAATT TATGTTACAG TTACCAAACCT GGGATATAGG AGTTTTGTTT	2400
ATAGTTGATT CTTTGAATGC TTGTGTGTTA GTCTTGTTTT GCTTTATGTT ATACGCACTA	2460
GGAGCGAATG TACAAATGAA CTTTGCTTGC GTTAGTTTCT TTTTACTCAT GATTGG	2516

(2) INFORMATION FOR SEQ ID NO: 249:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1364 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

CGGTGTTTTT TTGTAAATTT TCTAGCACTT GTATGGTAAA ATAGATACAG GTGTTCAATTA	60
AACTAGACTA AAAACCTATT TAAGCAGGCA AAATGAAGAA ATACCAACAA TTATTTAAGC	120
AAATCCAAGA AACCATTCAA AACGAGACTT ACGCTGTCGG AGATTTCCTT CCTAGCGAGC	180

1272

ACGACCTTAT GGAGCAATAT CAAGTGAGTC GTGATACCGT CCGAAAGGCC TGTCTCTCCT	240
CCAAGAGGAA GGATTGATCA AAAAGATAAG AGGGCAAGGT TCTCAAGTCG TCAAAGAAGA	300
AACCGTCAAT TTCCCTGTAT CCAACCTAAC CAGCTACCAA GAACTAGTTA AAGAACTTGG	360
ACTGCGCTCT AAAACCAACG TGGTCAGTCT GGACAAGATT ATTATTGATA AAAAATCCTC	420
ACTGATAACC GGTTCGCCAG AGTTTCGGAT GGTTCGGAAG GTGGTCCGCC AGCGTGTGGT	480
GGATGATCTG GTATCCGTTT TGGATACGGA CTATCTGGAT ATGGAACTCA TCCCAAATCT	540
CACTCGCCAA ATTGCTGAGC AGTCTATCTA TTCTTATATA GAAAATGGCC TCAAACCTCT	600
TATTGATTAT GCTCAGAAGG AAATCACCAT TGACCACTCA AGCGACCGAG ACAAGATTCT	660
CATGGACATT GGCAAAGACC CTTATGTCGT TTCGATTAAA TCAAAAGTCT ATCTCCAAGA	720
CGGACGCCAA TTTCAGTTTA CCGAAAGTCG CCATAAGTTA GAGAAATTTA GATTTGTAGA	780
TTTTGCAAAA CGCAAGAAAT AAAAGACTGA GACACCAGAT CTCAGCCTTT TTCGGCTCTA	840
TAATATTTGT AGTGGGTAAC CCCCCTATGG ATATTATGGA GCCTATTTTG TGTAGAAAAA	900
AAGTCCCATTA TGACCTATAA TGAAAAGCGA CAAAACAACT CATTAGAAAG ATTCATATGG	960
AACAATTACA TTTTATCACA AAACTGCTCG ATATTAAAGA CCCAAACATC AAGATTCTAG	1020
ACATCATCAA TATGGATACC CACAAAGAAA TTATCGCTAA GCTGGATTAT GAGGCTCCAT	1080
CTTGCCCTGA TTGTGGAAGT CTAATGAAGA AATATGACTT TCAAAAACCG TCTAAGATCC	1140
CTTACCTCGA AACAACTGGT ATGCCTACTA GAATTCCTCT TAGAAAGCGT CGTTTCAAGT	1200
GCTATCATTG TTCTAAAATG ATGGTCGCTG AAATTCTTAT CGTCAAGAAG AATCATCAAA	1260
TTCTCTGTAT TATCAACCAA AAAATTGCGC AAAAGTTGAT TGAGAAGATT TCTATGACCG	1320
ATATTGCTCA TCAGCTGGCC ATTTCAACTT CAACTGTTCAT TCGG	1364

## (2) INFORMATION FOR SEQ ID NO: 250:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1227 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

CCATGAAGAC CGCTTGGAAT TGGAATGGCA CAAGTCTTTG TTGAATGGTC TATTCCCATT	60
GACAATCGGT GGAGGAATTG GACAATCTCG TATGGCCATG TTCCTACTTC GCAAGAGACA	120
CATCGGAGAA GTGCAACAA GTGTTTGGCC TCAAGAAGTC CGCGATACTT ACGAAATAT	180
TTGTAGAGA ATCGAACCGC AAGGTTCCGT TTTCTTTCTC TTTTGTCTA TAATTTGGTA	240



1273

TAATAAACAG TATGAAAATC GTATCAGGAA TCTATGGGGG ACGTCCCCTC AAGACACTAG	300
AAGGCAAGAC GACAAGACCT ACTTCGGATA AGGTTAGGGG AGCCATTTTT AACATGATTG	360
GTCCTACTT TGAAGTGGGA CGAGTCTTGG ACCTTTATGC AGGTAGTGGT GGTTCATCTA	420
TCGAAGCAGT ATCGCGTGGC ATGTCCAGTG CTGTTTGGT GGAGCGAGAC CGTAAGCTCA	480
GACCATCGTG GCTGAAAATA TCCAGATGAC CAAGGAAGTT GGAAAATTTT AACTCCTCAA	540
GATGGATGCA GAAAGGGCAT TGGAACAGGT ATCTGGGGAA TTTGACCTCG TTTTCTTAGA	600
CCCTCCCTAT GCCAAGGAAC AAATCGTAGC AGATATTGAA AAAATGGCTG AGAGAGAGCT	660
TTTTCTGAA GATGTTATGG TTGTGTGCGA GACGGATAAA GCCGTTGAAC TTCCAGAAGA	720
AATTGCCTGT CTGGGTATCT GGAAGGAAAA GATTATGGA ATTAGTAAGG TGACAGTCTA	780
TGTCAGATAA GATTGGCTTA TTCACAGGCT CATTTGATCC GATGACAAAT GGGCATCTGG	840
ATATCATTGA ACGGGCGAGC AGACTTTTTG ATAAGCTTTA TGTGGGTATT TTTTTAATC	900
CCCACAAACA AGGATTTCTC CCTCTGAAA ATCGTAAACG GGGGTTAGAA AAGGCTGTGA	960
AACATTTGGG AAATGTTAAA GTCGTGCTT CTCATGATAA ATGGTGGTC GATGTCGCAA	1020
AAAGACTGGG GGCTACTTGC CTAGTGCAG GTTTGAGAAA TCGCTCGGAT TTGCAATATG	1080
AAGCCAGTTT TGATTACTAC AATCATCAGC TGTCTTCTGA TATAGAGACT ATTTATTTAC	1140
ATAGTCGACC TGAACATCTC TATATCAGTT CATCAGGCGT TAGAGAGCTT TTGAAGTTTG	1200
GTCAGGATAT TGCCTGCTAT GTTCCCG	1227

(2) INFORMATION FOR SEQ ID NO: 251:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3652 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

CCGGTCAAGT TAAAAACGCT ATTTCTTCCC ATTTTATTTA TTTTITAGGA GTGGTAACGT	60
ATCAAAATAG CCCAAGCGTT CTCACCCGTG TGAGTTTGAA TAATGGAACC CGTTTCCAAA	120
ACAGAAATTG GCTTTTCAAC ATAAGCTTGT AAGCTTTCTT TCATCTCTTT TGCCCAATCA	180
TCACTACCAG AATATGAAAT TCCAATCTCT GCTACAGCAC GTTCAGAAAG CGATGTTATC	240
AACTCATCTA ACCATTTTTT AAATGTTTGA GTTCCACGAC CTTTAACCAT TGGCTGCAAT	300
TCATGGTCTT TCATTTGCAT GACAGCACGG ATATTGAGAA GAGAGCTCAA CAAGCCAGTT	360

1274

ACACGGCTAA TTCGTCCACC TTTGACAAGA TTTTCCAAAG TTGAAACACC AATATAAAGC	420
TCTGTATGGT TTTTAACCTC TTCTACATGA GATAAAATTG CCTCCATATC TTTACCTTCT	480
TGAGCTAACT TCGCAGCCTC AACAACTTGG AATTTTCAGGG CTGCGTCAGT GAAGGAACTA	540
TCAACAACAG TCACATCTGC AGTAGATAGG CTAGCACCTT GGCGTGCTGC TTCTACCGTA	600
CCCGAAAGAG CATGGGACAT ATGAATAGCA AGAATCTGGC CACCATCTTT GCATAGGTCT	660
TCAAAAATCT CAGCAAAGAC ACCTACAGGT GGCTGACTTG TTTTCGGAAG ATTCTTACTT	720
TCTTGCATCA ACTGAAGAAA TTTACCTTCT TCTTTCAAAT CCGCATCAGA ATAAACAACA	780
TTATCAATCA TTACAGATAA TGGAACAATT GTAATATCTA ATTGCTTTAC TAGTTCAGGT	840
TCAATAGTAA CAGATGAATC GGTTACAATC TTAATTTTTG TCATAGTATC AATCTTTCTA	900
TTTTAGGATT CAGATTGGTT TCCTTACTTC TAATTATATC AAAAAAAGA TTA AAAAATCC	960
TAATGGAGTC AATCAAATTT TCCGTAAAAT TTGATATAAT CAACTTATAA GAAAAGAGGT	1020
GTCCTATGAT TAAAAAAATT TACCCCATTT TTACCATTTT ACTAGGTGCT GCTATTTATG	1080
CTTTGGACT GACTTATTTT GTAGTTCCCC ATCATCTCTT TGAAGGAGGG GCGACAGGCA	1140
TTACCCTCAT CACCTTTTAT CTTTMTAAAA TCCCTGTTTC CCTCATGAAC CTGCTGATTA	1200
ATATTCCCTT TTTTATCCTA GCTTGAAGA TTTTGGAGC CAAATCCCTC TATTCTAGTT	1260
TACTAGGAAC CTTAGCTTTG TCCGGCTGGT TAGCTTTTTT TGAGCATATT CCCCTTCATA	1320
TTGATCTTCA AGGTGATTTA CTAATCACAG CCCTTATAGC GGAATCCTA TTGGGAATTG	1380
GCCTTGAAT TATTTTAAAT GCTGGAGGTA CAACTGGCGG AACTGATATT CTAGCTCGTA	1440
TTCTCAACAA ATACACTCAT ATATCCATAG GAAACTGCT CTTTATCTTA GATTTTGTGA	1500
TTCTCATGTT GATTCTCCTA ATCTTCAAGG ATTTGAGATT GGTTCCTAC ACGCTTTTGT	1560
TTGATTTTAT TGTTCCTCGT GTTATTGATT TGATTGGTGA AGGAGGATAT GCCGGCAAAG	1620
GCTTTATGAT TATCACAAAA CGTCCTGACC AACTTGCTAA GGCGATTAAT GATGACCTCG	1680
GAAGAGGTGT TACTTTTATT TCTGGTCAAG GCTACTATAG TAAAGAAAAT TTGAAAATCA	1740
TCTACTGTAT TGTCGGAAGA AATGAAATTG TGAAAACGAA GGAAATGATT CATCGAATCG	1800
ATCCTCAAGC CTTTATAACT ATTACAGAAG CCCATGAAAT CCTAGGAGAA GGCTTCACCT	1860
TTGAAAAAGA ATAAAAAGAG GTAATGTCGT GACCTCAAAA GTTAGACTAA ATCATCTATC	1920
TTTTGGGTTA CAGACAACCT CTTTTTTATT TTATTTACTC AAGCTCTTAA GACCAATTCC	1980
GAGTTACTTC TTCATCAGCC TTAACTGAT CCACTAATTG GTCAACTGAG TCAAATTTGG	2040
TCATATCTCG AATGCGATCA AGCCAATAAA CCATGACGGT TTCCCCATAA ATATCTTGAT	2100
TAAAATCAAA AATATTGACT TCAAAACGTG CTTCTTCTCC ATCAAAGGTC ACATTTTTC	2160

1275

CGACACTAGC CATAGCACGA TACTTCTGTC TTTGAATCTC AACATCAACA ACATAAACGC	2220
CATCTGCTGG CATATAAGTA CGGTCTAAAA GCACTAAATT CGCTGTCGGA TAACCAATTG	2280
TACGACCACG AGCATTACCA TGAACCACCA TACCTCTTGA TGGAAAGCGGT GCCCCCAAAA	2340
GTTTTCTGCG TTCTTTCACA TTTCCATCTA AAATAGCTTG ACGGATACGA GTTGAACATA	2400
TCTTTCCTTT CTCATCTTCT ACAGGTGGAA CAATGATAAC TTCTCCATCA AAGTAATTCT	2460
TTAAATCTTC TGCTGTTTTT TTGTCAGAAC CAAATGTATA ATCAAAACCT GCAACAATAA	2520
TTTTGGCATT CATAGCCTTG ATATAAGTTG CAAAGAATTC TTGTGCAGTG AGACTAGCGA	2580
ATTGACTACT AAAATCAAGG AGATATAATT CTTCTACACC TTCGCGCTTT AATTTTCTTT	2640
CACGTTCAGC AGGGTTCAAA ATATGCAAAA ACAATCTGG ATGATAAGGC TCTAAAGCGA	2700
TCTTTGGAGA TTCATTAAAG GTCATAACGA CGATAGGCAA CAAATCCTTT CTCGCAGCCT	2760
TGTTGGCAAC ACGAAATAAT TCTTGATGCC CCTTATGTAT GCCATCAAAA TAGCCGAGAA	2820
CAACGACTGA ATCAGATGGT GTGCCAATAT CTTTTGGTT TTTTATAGGA ATAGTAATAA	2880
TCATAAATA ATTATATCAT AGCGATAGCT ATTTCTGGAA CAGAAAATCT GAAATGTTGT	2940
TTTTTTCACA TGAAGGTAC CTGTTTCAA AAAGCACTTT ATTCTATCGT TGCTTAATA	3000
TGAACTTTGC AATATTCTTC TCAAAAACCT GTAGGACATC TTCAAAATTT TGCAAGGAGT	3060
GATTAGACTT GTTCGGTAAC CATAAAGTG CATACTATGC TTATGTATGA AAAAGCAATG	3120
CAACTAACTC CTGAGAACTT TAAATTAATA ATTGGTGCCG AAAAGGTAGA ATTTAGAATC	3180
GAGGTACACC TATGGCTGTA AAATTTACAA AATGAGACAA CTTGGGCAAG ATGTTTGAAG	3240
AATTTCTTAA ACTCCCTGAT TTGAAGCAAG TCACTTTCCC TAATGACAAA GAAAAAGCC	3300
AAAACAGCAA AGAAAACTA GATGACTGCT TTCCAACAAC TCCCATCTAG TGTGCTTCAG	3360
ACTGGGCTAT TTTTCTCTCC ATCTGTTAGC TTGGATTCTC AGACCGTTTC AGCTAAAGAA	3420
TATCTTTTCC CTTATCAGAA GGAACGGCTC AAGCCATTCA GACAAGTGAA GGGACGACAA	3480
GCCAATATTT GAAACCAGAT AGCAGTTCTT ATAGTCAATT GAAATAAAAT CTGAAGAAAT	3540
CGAGTAGGAA ACTCATATCA ATGTTTAACA GTGTTCTATT CCAGATTCAT ACTCAATGAW	3600
AATTAAAGTG CAAACTAGGA AGTTAGCCGC AGGTGATACT TTGGGTACGG CA	3652

(2) INFORMATION FOR SEQ ID NO: 252:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 743 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1276

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

GTACCGTGGT GCCAAAGTAC AGCAAGGTTG GCTTTTTGAC AAACAATACC AATCTTGGTT	60
TTACATCAAA GAAAATGGAA ACTATGCTGA TAAAGAATGG ATTTTCGAGA ATGGTCACTA	120
TTATTATCTA AAATCCGGTG GCTACATGGC AGCCAATGAA TGGATTGGG ATAAGGAATC	180
TTGGTTTTAT CTCAAATTG ATGGGAAAAT GGCTGAAAA GAATGGGTCT ACGATTCTCA	240
TAGTCAAGCT TGGTACTACT TCAAATCCGG TGGTTACATG ACAGCCAATG AATGGATTG	300
GGATAAGGAA TCTTGGTTTT ATCTCAAATC TGATGGGAAA ATAGCTGAAA AAGAATGGGT	360
CTACGATTCT CATAGTCAAG CTTGGTACTA CTTCAAATCC GGTGGTTACA TGACAGCCAA	420
TGAATGGATT TGGGATAAGG AATCTTGGTT TTACCTCAAA TCTGATGGGA AAATAGCTGA	480
AAAAGAATGG GTCTACGATT CTCATAGTCA AGCTTGGTAC TACTTCAAAT CTGGTGGCTA	540
CATGGCGAAA AATGAGACAG TAGATGGTTA TCAGCTTGA AGCGATGGTA AATGGCTTGG	600
AGGAAAACT ACAAATGAAA ATGCTGCTTA CTATCAAGTA GTGCCTGTTA CAGCCAATGT	660
TTATGATTCA GATGGTGAAA AGCTTTCCTA TATATCGCAA AGTAGTGTCG TATGGCTAGA	720
TAAGGATAGA AAAAGTGATG ACA	743

## (2) INFORMATION FOR SEQ ID NO: 253:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4010 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

TTTGGTTGA TGATACGAGG GATTGGTGA TTCTTCTTGA CGATAGAAGT TTCAGCGACC	60
ATCATTTTTG AACAGTGATA GCACTTGAAT CGACGCTTTC TAAGGAGAAT TCTAGTAGGC	120
ATACCAGTCG TTTCAAGATA AGGAATTTTA GAAGGTTTTT GAAAGTCATA TTTCTTCAAT	180
TGGTTTCCGC ACTCAGGGCA AGATGGGGCG TCGTAGTCCA GTTTGGCGAT GATTTCCTTG	240
TGTGTATCCT TATTGATGAT GTCTAAAATC TGGATATTAG GGTCTTTAAT GTCTAGTAAT	300
TTTGTGATAA AATGTAATTG TTCCATATGA TTCTTCTTAA TGAGTTGTTT TGTCGCTTTT	360
CATTATAGGT CATATGGGAC TTTTCTTCTA CAATAAAATA GGCTCCATAA TATCTATAGT	420
GGATTTACCC ACTACAAATA TTATAGAACC GAATTAATTT AATTAGAGAG CCAACTTTCT	480
AATATAGTAA TCGCGTCATA ACAAGGTATC TATCATTCAT GGAGTTCCTC CTGTATACTA	540

1277

TTAGTAAAGT AAAACTATTG GAGGATATTT TAATGCCACA ACCTATTGTT CCTGTAGAGA	600
TTCCACAATC TCGTCGTTTT GATTCTAAAA AGAGAAATGA TATTCTGCTT AAAATTCGTA	660
TTGGCAAGCT TGAAGTAAGT TTTTTTCAAT CTCTCAATCT CGAAATGGTA GAACAGCTTT	720
TGGATAAAGT GTTGCTCTAT GACAATTCAT CTATCTAGCC TAGGGCAGGT CTATCTCGTA	780
TGTGGGAAAA CGGATATGAG GCAAGGCATT GATTCAATGG CTTATCTGGT TAAAACCCAC	840
TTTGAATTAG ATCCTTTCTC CGGTCAAGTT TTTCTCTTTT GTGGTGGACG TAAAGACCGC	900
TTTAAAGCCC TTTACTGGGA TGGTCAAGGA TTTTGGCTAC TATATAAAGC CTTTGAGAAC	960
GGAAACTGA CTTGGCCCAG TACAGAAAAG GATGTCAAAG CTCTCACACC TGAACAAGTA	1020
GATTGGCTTA TGAAGGGCTT TTCTATCACT CCAAAAATAA ATTTATCAGA AAGTCGTGAT	1080
TTCTATTGAA ATGAGGACTT TCTTTTAGT TATAATAAAG TTAGGAAATA AGGAGAGGAA	1140
GCCCATGGAA GAAGATTGAA AATCATTCAA CAACAGAGTG CTACAATTGA TAGTCTCACC	1200
AATGAACCTG CCTTCTTCG TGAACAAGTG GCTTATCTAA CGCAAAAGCT CTATGGAAAA	1260
TCCTCTGAGA AAAGTGTTG CCCATCTGGA CAACTCAGTC TTTTGAAGA GGAACAAAAT	1320
ATGGAAGAAG ACTCTGACTT ACCCAGTTGA AAGAGAAGAA ATCACCTATA AACGTAAGAA	1380
AGCTAAAGGG AAACGTCAAG CTCTTCTTGC CCAATTTGAT TCAGAAGAAG TTCATCATCA	1440
AGTAGAAGAG AGCATTTGCC CTGATTGTCA GGGAGATCTA AAAGAGATTG GAGCAACCCT	1500
TCAACGACAA GAATTAGTCT TTATTCTTGC GCAATTAATA CGAATAGATC ATATCCAACA	1560
CGCTTATAAG TGCCAAGCAT GCAGTGATAA AAATCCGAGT GATAAAATCG TGAAGCTCC	1620
TATTCTAAA GCCCCTTTGG CGCATAGCCT TGGCTCAGCT TCTATTATCG CTCACACCAT	1680
CCATCAGAAG TTTAATCTGA AGGTACCCAA TTATCGCCAA GAAGAAGATT GGGCTAAGAT	1740
GGGTTTACCA ATCACACGTA AGGAAATTGC TAATTGCCAT ATCAAGGCCA GTCAATACTA	1800
TTTGGAGCCC CTTTATAATC TTTTACGAGA AAAGTTGTTA GAACAAGCTC TTCTTCATGC	1860
GGATGAAACC TCTTATCGGG TTCTAGAGAG TGATAGTCAG TTGCCTTACT ATTGGACTTT	1920
TTTGTCTGGG AAAGCTGAGA ATCAAGCAAT CACGCTGTAC CACCATGATC AGCGTCGGAG	1980
TGGTTTAGTA GTACAAGAAT TCCTAGGAGA TTATTCTGGC TATGTTTATT GTGACATGTT	2040
GCGGCAGTAA CTTAGGACTT TAGTCCTCTA GTTCTGCCTA TGCGATAGCA GTCCAAGGTT	2100
TAGGAGTAAG GCGACGCTAA GCTTGGTAAA CTGCGAACAG CTAGAAGCTT ATCGTCAACT	2160
GGAAGAAGCT GCACTTGTG GATGTTGGGC GCATGTGAGA AGGAAGTTTT TTGAAGTGCC	2220
CCCCAAGCAA GCAGATAAAT CATCCTTAGG AGCTAAAGGT TTAGCTTATT GTGATCAGTT	2280

1278

ATTTTCCTTG GAAAGAGACT GGGAGGCTTT GCCAGCTGAT GAACGACTAC AGAAACGTCA	2340
AGAACATCTC CAGCCCCTAA TGAAGACTT CTTTGCTTGG TGCCGCCGTC AGTCAGTTTT	2400
AGCAGGTTCA AAAGTAGGAA GGGCAATTGA ATACAGCCTC AAGTATGAAG AAACCTTTAA	2460
GACTATTTTG AAAGACGGAC ATCTGGTCCT TTCCAATAAT CTAGCTGAAC GCGCCATTAA	2520
ATCATTGGTT ATGGGACGGA GTAAAAGAGT CCAGTGGACT CTTTATAGCTT GAGCTCAGTT	2580
TAAAAAAGCG AGGGTGGTTA TTTTCTCAA GTTTTGAAGG AGCTAAAGCA AGAGCTATTG	2640
TTATGAGCTT GTTGGAACA GCTAAACGTC ATCAATTATA GTGCGTTGAA TCTATAACAG	2700
TACGCATCGA CTGCTAAAC ATTTCTATAA ATCAATTTTC CTTTCCTAAT CGATTTGTTT	2760
ATATCTTATT TCAATCCATT ATAAATAGCG AGAAATATCT ATCCTATCTT CTAGAATGTC	2820
TTCCAAACGA GGAAGCTCTC GTAAACAAAG AGGTTTTAGA GGCCTATTTA CCGTGGACTA	2880
AAGTTGTACA AGAAAAGTGC AAATAAGAAA TCTCCAGATT AGGAACTATC CGTGAGTTCT	2940
CTAGTCTGGA GATTTTTCAA TAGACTTCGT TATTTGGACGG TTACAATTTA TTATATGAAA	3000
ATCCCATATT ATTCTCCAAT TCTATATTTT ACCTTTCTAA ATGTATAGAT TAACTACCTA	3060
ATTATAGCAT ATAACGCAGA TTCCTTTCAA TCGTATGATT TACTGCATTA AATTAAGTAA	3120
AAAAATAAAG GCAGTCCGAA GACTGCCGAT ATTTATCTCT CATCTCTTTA ATTATGGTAA	3180
GTAATAAAT AATTTCCCTA AAGATATGGA AATTATTAAT ACTATAAATA CATATTATAA	3240
AGTTTATAAA TACTGTAAAA ATCCTGAAGT TAATTTTCTA ATAAATATCA ATATGTGTAA	3300
GTATCTTTTA AATTTTTAGA CAATTTACTA GTTCTATAGA CATGTTTAAC AGACTCTATT	3360
TTACAATTCA AAAATTTTCA CTGCCACTTC ATTTAAAAAT TCTATATCAT GGGAAACAAT	3420
AAAAATTATT TTATCCATGG TTTTATACTT ATTAATCAGT TCAGATATTT TTATCATATT	3480
GGAATAATCC ATACCACTTG AAGGTTTCGTC AAAAAAGACA AATGGAGAAT TCTTGACAT	3540
AACAGATGCT ATTGCAAGCC TTTGCTTTTG CCTCCTGAT AAACATCATG GATGCCTTTC	3600
AATAAATTCG TCCAGGCATA AATCTTTTAA CCCAAATCAT TCATACCTCT CTCAACTAGA	3660
TGTAACTTAC AAAACCCCTG ACCTCATGAG CCACTTTCTT CCTCCTCATG AGGTCAGTTT	3720
TACTTTCTGC TGTCCAGTA TCGTTTTTCC TCGCTAGATT TCCTCAAAG GGCAGACTCC	3780
TCCCTTGGTT CGTCACACGA TTTTTCATC TCGACTGTTT TTTAATGCAT CATTAAACGAC	3840
GCTTTTCTTC TAGGTGGTTC ATAAGGAACA GGAAGATTCA GGTTGACTTT TCTAATCCTA	3900
GAATAAAGTG CTGAAAACAA TTCGGAATAG GCATAGAGAC TAGACAATTT GAGGAGCTGC	3960
TTGCGTCTG TTCGAACACA TTTTCCACC ACGTGAAGAA AAAGATGGCG	4010

(2) INFORMATION FOR SEQ ID NO: 254:

1279

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2789 base pairs -  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

ATGCATCCGT TTGTCAAGCC TAAATTGTAA TTTTTCCTCA TTTAAACAG AAAACCCAG	60
GAAAATGACA TAAAAATATC ATTCCTAGGC CTATTTATGC TATTTCTCTC TGAAAAATAT	120
GAGTATTCAG TCGGTCAAAT GAAGCTGAAC GAACTCATTT TCCCTCGCCT AATTCATGA	180
TTCGATGACA TTGTTGGGCT ACATAAGCAT CGTGGGTAC GATAATGACT GTTTTCCCT	240
CTCGATTCAT CTCTAAGAGA AACTTCAAGA CCAAATCTCT ATTTTCAGGA TCCAGAGAAC	300
CTGTCGGTTC ATCGGCTAAA ATCAGCTGGC TGGGTTTAA GATGGCTCTA GCAACTGCAA	360
TTGTTGTTG TTCGCCCCA GACAACTCGG AGACCTTTG ATGCAAAGTA GCTGATAAAC	420
CTACTCTCTC TAAATCTCT TCCACCTTTT TGAGCTTGTC TTTCTTAGGC AATTCACAT	480
ATTTCAGCGC CACATGAGAT TGTACTCGAC CGTTTCATCA TCAATCAGGG CAAATTTTG	540
AAACAGATAA GAGATATGTT CACGGATTAT TGTTCGCAC TTAGCAGAAT TAACCGCTAG	600
ATTTGTCTGA CCAAAATCT CATACCGTCC GCTATAATCA CCATCTATCA AACCCAATAA	660
ATTTAACAAG GTCGACTTCC CACTACCACT CTTACCAACA ATAGCTACCA AATCCCCCTG	720
ATCAATCCTG AGAGATAAGT TATCCAAAT CACTTTTCCC CCAATGGTTT TGGTAATATT	780
TTTCAACTCA ATCATAAGAT GCGCCCTTTC AATAACTCTA CTAGACTTCT TTTCTCCATC	840
CTAGAAGCTA AGCCTAGCAC AAATAGTATA TCCAGACATG TAAACCTGC AAACAGTAGA	900
AGTGGTAAGA ACGCATGGGC AAAGAAAATC AAGACTAGAA GAGGGAAAT ATAGCCCAGC	960
AAGAGCAGAA CGAGGAGAGG ACGGTAGCGA TCGACCAGTT TCCACCCCAT AAATTCTTGT	1020
GTAATGATAT CCCTGCGCTT CAATAAGAAA GTTGTACTA GTAAGAAGTA GGAAATCATC	1080
ATGCTAAGGA GACCAACAA AGCAAAGAGT AGGTTAAAT TCCGAACAGC ATCTCGATAA	1140
GAATCCACTT TCTCTTGTG AATGGCTTGA ATAGATGAAA ATTTTAAATA ATTTCCATCT	1200
GACAATTTCT CAACTAACTC TGTAACTCT TTTTGATGTT GAACCGTATT TTCAATTTTA	1260
ATCGGATTAT TTAAGCCAGT TGTGACAGG GAGGCTTCT CATCCACAT CATATCAGAA	1320
TCATTGACCA AGCTAATAAT TGGATTGGAG AGATTTCCT TTCGCTTATC ACTATATGGG	1380
AAAAATGACC AATCTCCTTC ATAATAGGCA ATCTCGACAT CCATCTCCTC TATCGTTCGT	1440

1280

TTTGTGCTGCT CTTCACTATT CATCGAATGA AAGGCAATTA ACTTCCCCAA GAGCTGATTT	1500
TTATCTTCTT CACCTTTCGT ACTTGCTGGC ATCAAAATAA CTTTTTTAAT ACCGGTATTT	1560
GGTAGCTTGA ATCCCTTGCT CTTTAGAAAA TTGCGATTGG CATAGTAAAC ATCCACCGTA	1620
TCTGTAACT GATATTGCTG AATCTGTTCT GATTGGACAA AATTTTTTAC AGGAAGACTG	1680
CTACTCTGCA CATAGCCCCG CTGCGTTTTT TCTACCAAAT CCTGATAAAA TCGATAGAAA	1740
TAATCTGTAG ATTTCCCTGA CCCTGCTAGC TCTTCTTGCC ACAGATTATC ATTGAGTTTG	1800
AAGGTTTCTA AGGTCAGGTA ATTACCTTGA CTTACCCACT GTTGCTGATA AGCAAGTTCT	1860
TTGTTTTCTT GTTCTAACT TCTGCCACC CCAATCAGTA AGGCCGTCAG TAAATAGTT	1920
GTCCCTATTT TCATCACATA ATTGAAGATA AGACCAAATT TGAAAGATGA AAAACCTTTC	1980
AGCAGAGAGC TGATTGTCAT TTTTGGATT AAGAGGTAAG TCAACCAACT GATAAAGAGA	2040
TAAAGCTGCA ACAGCAAAAA ATGAGACAAC CACAGCATAG GAAACAAATC TTTTGGCTTA	2100
TAATCAAGCA AGAAAAACAC GCCTAGATTG ATCACAAGAG CCCACCTAG GAGGAGGTAA	2160
AGGTTGCCTT TTACAACATC AGCTAAAACA GCCCTATCTT GAAAACCAAG TAATTTTGT	2220
ACCCCACTC TTTTCATCTC CATCATCGGT TGATACACTG TCACTAACAC AAGAAGCAAA	2280
ATAGCCAAGA CAAAAACAAT GGCAGATAAA AGCAAATCTC GATTTATGAC TTCCACTGCA	2340
CTTTTGTAGG TCGGCTCTAG CAAGGTAGCC TGGTCTATCT TGAAAAAATC GCTCCATTT	2400
TGTACAATCC TATCCTTGTC CATCTCTTGT GTAGAAGTTA TCGTATAGCG ACCATTTAAA	2460
CTACGAGATG TATCCTTGAT ATAGGTTTGA AAAGTCATAA GCTGAATAGG TTTGGCTTTT	2520
AGAAAGGTCG GAATCGTACC AAGTTTATTG GAAATTTCTT TATTACTATA GACTCCTTCA	2580
CCATCTGTGG TAAAATCAAG AGAAGAAATC CCAAATCTT GGTAGGGGAA GGTATCTTTA	2640
TCAAAAACAC CAGACTTGAC CACCTCATCA CCACTGTCTG TTTTGATGAT GGAGACTTTA	2700
TACTCCTTTG ATACATCCTC AAAAAATCGA AGAACAGACG CTGCAGGTTC GTTAATATCT	2760
TTCAAATACA AATCAAAGA ATCTACAGG	2789

(2) INFORMATION FOR SEQ ID NO: 255:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2495 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

CTGCGAATTT TATTAAAGAT AATGTGTTAA TTACAGCGGC TCACAACTAC TACAGACATG 60



1281

ACTATGGGAA AGAAGCGGAT GATATTTATG TTCTTCCGGC TGTTAGTCCA AGTCAAGAAC	120
CATTTGGAAA GATCAAAGTA AAGGAAGTTC GTTATTTGAA GGAATTTAGA AATTTAAATT	180
CTAAGGATGC AAGGGAATAT GACTTGGCTT TATTAATTCT AGAAGAGCCC ATTGGTGCAA	240
AATTAGGGAC TTTGGGTCTT CCTACTAGTC AAAAAAATTT GACAGGAATA ACTGTGACTA	300
TCACAGGCTA TCCATCATAT AATTTTAAAA TTCATCAAAT GTATACAGAT AAGAAACAAG	360
TTTTAAGTGA TGATGGCATG TTCTTGGAAT ACCAAGTTGA TACTTTAGAG GGGTCTAGTG	420
GATCTACAGT TTATGATGCT AGTCACCGTG TAGTAGGAGT GCATACTTTA GGAGATGGAG	480
CTAATCAAAT TAACAGTGCA GTTAAATTAA ATGAACGAAA TTTGCCATTT ATTTAWTCGG	540
TTCTTAAAGG TTAATCTCTT GAAGGATGGA AGAAAAATAA TGGTAGTTGG TACCATTATA	600
GACAACATGA TAAACAAACG GGTGGGCAGG AGATAAATGA TACCTGGTAT TATTTAGACA	660
GTTCCGGTAA GATGCTTACA GATTGGCAAA AAGTCCATGG AAAATGGTAT TATCTCAATT	720
CAAATGGAGC AATGGTTACA GGTAGCCAAA CTATCGATGG TAAAGTTTAT AACTTCGCTT	780
CATCTGGTGA GTGGATTAA TGTTGGAGGA TATATAAAAT GAAGCTTTTG AAAAAATGA	840
TGCAAATCGC ACTAGCCACA TTTTCTTCG GTTTGTTAGC GACAAATACA GTATTGCGAG	900
ATGATTCTGA AGGATGGCAG TTTGTCCAAG AAAATGGTAG AACCTACTAC AAAAAAGGGG	960
ATCTAAAAGA AACCTACTGG AGAGTGATAG ATGGGAAGTA CTATTATTTT GATCCTTTAT	1020
CCGGAGAGAT GGTGTGCGGC TGGCAATATA TACCTGCTCC ACACAAGGGG GTTACGATG	1080
GTCCTTCTCC AAGAATAGAG ATTGCTCTTA GACCAGATTG GTTTTATTTT GGTCAAGATG	1140
GTGTATTACA AGAATTTGTT GGCAAGCAAG TTTTAGAAGC AAAAACTGCT ACGAATACCA	1200
ACAAACATCA TGGGGAAGAA TATGATAGCC AAGCAGAGAA ACGAGTCTAT TATTTTGAAG	1260
ATCAGCGTAG TTATCATACT TTAAAACTG GTTGGATTGA TGAAGAGGGT CATTTGGTATT	1320
ATTTACAGAA GGATGGTGGC TTTGATTCGC GCATCAACAG ATTGACGGTT GGAGAGCTAG	1380
CACGTGGTTG GGTTAAGGAT TACCCTCTTA CGTATGATGA AGAGAACTA AAAGCAGCTC	1440
CATGGTACTA TCTAAATCCA GCAACTGGCA TTATGCAAAC AGGTTGGCAA TATCTAGGTA	1500
ATAGATGGTA CTACCTCCAT TCGTCAGGAG CTATGGCAAC TGGCTGGTAT AAGGAAGGCT	1560
CAACTTGGTA CTATCTAGAT GCTGAAAATG GTGATATGAG AACTGGCTGG CAAAACCTTG	1620
GGAACAAATG GTACTATCTC CGTTCATCAG GAGCTATGGC AACTGGTTGG TATCAGGAAA	1680
GTTCGACTTG GTACTATCTA AATGCAAGTA ATGCAGATAT GAAAACAGGC TGGTTCCAAG	1740
TCAATGGTAA CTGGTACTAT GCCTATGATT CAGGTGCTTT AGCTGTTAAT ACCACAGTAG	1800

1282

GTGGTTACTA CTAAACTAT AATGGTGAAT GGGTTAAGTA ATGAAGGCTA ATTGTAAACT	1860
GTGATGGATA CTAACTTTG TATAATAGGT GGATAAAAGT CTTCAACAATC AAAAAACGCA	1920
TAGTATCAAG GTTTTCTGT ACTGCCCTCA AACAGTTAGA CAATTAATTT ATCCGAAGGA	1980
TTTAGTTCTG TATTGCACAG GGCTAAGTCC TTTAGTTTT ACCTTAATTC GTTTATTGTT	2040
GTAGTAATCA ATATAGTCTA TAATGGCTTG TTCCAATTGC TTAGCGACT GAAACGACTT	2100
CTCATAACCG TAAACATTT CCGATTTTCAAG AATCCCAAAG AAGGACTCCA TCATACTATT	2160
GTCTGGGCTG TTTCCCTTAC GTGACATGGA TGCTTGAATT CCCTTACTCT CTAGGAACCG	2220
ATGATAAGAA TCGTGTGGT ATTGCCAGCC TTGGTCACTA TGGAGAATCG TATTCTCGTA	2280
GTGCTTCTCT GTGAATGCCT GTTCCAACAT TGTGTGACT TGTCTAAGT TGGGTGAAGT	2340
TGAAAGATTA TAGGCGATAA TTTCGCTATT AAAGCCATCT AAAACTGGTG ATAAGTAAAG	2400
CTTTTGAGTA CTTGCTGGAA TGGCAAATTC TGTCACATCT GTGTAGCACT TTTCCATTGT	2460
TTTAGAGCCT TCAAATTGGC CTGAATGAG ATTCG	2495

## (2) INFORMATION FOR SEQ ID NO: 256:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 870 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

TACCACCGTA TTCATCCAGC AAGATTGCCA TTTGTCTTTG GGTATTTGCG AGTTCTTTTA	60
GCAAGTCATC CACAAAAATA GTTTCAGGTA CAAAAAGTGG ATCTTGTAAG ATTCTCTTCC	120
AAACAATATT GTCAAAACCG TCCACAAAGC CTGCCTTAAG GAGACTCTTG GTGTGAATGA	180
TTCCAATTAC ATTGTCTTTC TCCCCATCAT AAACCGGGAT ACGAGAATAA TTTTGTTTTA	240
AAATACTTTG GATAATGGCT TGAATATCAT CCTGAATATC CACCATAAAG GCATCCGTTC	300
GAGGAACCAT AACCTCTCGT GCCATCAGTT CATCGAGCGA AAAGACACCT TGTAGCATCT	360
CAATCTCATC AGCATCCAAT GTTCTTTCAC TATTTGTCAG CATATAGGCA ATTTTCATCAC	420
GGGTCATCTT TTCATCCGCA TCATCGAATG ACATAGGAGT CAAATGGCTC AAGAAATTGG	480
TCGAAGCAGC TAAAGCCAA ACAAAGGAC TGAATAGTTT TCCGATCCCA ATGATAATCG	540
GCGCTGTACG AATTGCCAAG GCATCCTTTA GATTAAAGAG GATTCTCTTA GGATATAATT	600
CCCCAAAAAC GATGGAAATA TAGGTCAAAA ATGCCAAGGA TAGAAAAGTT GCCACGGCTT	660
GTGCTGTTTC GCCATTCCCA AGCCAAGAGG CAATCACACG TCCTAGAGTA TCAGTTAAAC	720

1283

TCGCCCCCTGA TAAGATTGTA ATCAGGGTGA TTCCTACCTG GATGGTTGAT AAAAAGTGGT	780
TAGGATTTTC TAGTACCTTC AGCAGGCGGA TGTAGCGTCT GTCTCCTTCT TCCGCTTTT	840
GTTCAACTCG GGCACGATTA AGAGAAACGG	870

(2) INFORMATION FOR SEQ ID NO: 257:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1245 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

CGTCCCAGA AGCCCGCATT CTCATCGCCA ATGTCGTGAT TGATTGGCC CTTTCTCCAA	60
AATCCAATC AGCCTATGTA GCTATGGATA AGGCACTTGC TGACCTCAA ACATCAGGGC	120
ACTTGCCTAT TCCGCGACAC CTGCGTGATG GGCACCTACAG TGAAGCAAG GAACTGGGGA	180
ATGCCCAAGA CTATCTCTAT CCACACAAT ATCCTGGAAA TTGGGTCAAG CAAGACTATC	240
TGCCAGAAAA AATTCGTAAT CATCACTATT TCCAAGCAGA AGATACTGGT AAATATGAAC	300
GGGCTTTGGC TCAAAGAAAG GAAGCTATCG ACCGTTTGGC AAAAATCTGA AATCCTTTTC	360
AAAAAATTGC ACTTTCCTCT TGATTTTTTT TGAAGAGTG GTATCATATA AATATAGAAA	420
CGCTGTGGTG TACGACTTCA CACTTAAGTG TTGACCGACT ATTTTTTGTA TTATTAGGGA	480
AACAAAAGTC TTCTAACAGC ATGTAGGCCG TCTCACACGG AACAGCTTC AGTTAGAGCG	540
AGTTGCCCAC CTGCTTAATT GCGCGGTTT AATACAAACC GTGAAGTTTC GGCACCAATA	600
CAGCTTTTTT CTTTGCCTCC TTAGCTCAGC TGGCAGAGCA GCGGACTCTT AATCCCTGGG	660
TCACAGGTTT GATCCCTGTA GGGGGCATAT AAATACAACA GGAAAAGCCT TATAATATAG	720
GGCTTTTTTT GCTTTCCTTT TAAAAATTGT CGTGCAATTT GCCGTGTTTT TACAACAAAC	780
TTTTACAGC CATAAACTCC TCACTAATTT TTTCCTCCAA GGTATGCCCA TAAACGTCAA	840
TCAACATGGA GATATCTTTA TGTCTAAAA TTGGGCTCTT TGTCAACTGT AGTGGGTGTA	900
AGTCAGCTAA GCTCGAGAAA GGACAAATTT TGTCCTTTCT TTTTGATAT TCAGAGCGAT	960
AAAAATCCGT TTTTGAAGT TTTCAAAGT CCGAAAACCA AAGGCATTGC GCTTGATAAG	1020
TTTGATGAGA TTATTGGTCG CTTCCAATTT GGCCTTAGAA TAGTGTAGTT GAAGGGCGTT	1080
GACGATTTTC TCTTGTCTCT TTAGAAAGGT TTTAAAGACA GTCTGAAAA GAGGAGGAAC	1140
CTGCTTTAGA TTGTCTCAA TGAGTCCGAA AAATTTCTCC GGTGCCTTAT TCTGAAAGTG	1200

1284  
AAACAGCAAG AGTTGATAGA GCTGATAGTG ATGTTTCAAG TCTTG

1245

(2) INFORMATION FOR SEQ ID NO: 258:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1684 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

ATGCCTATGT AACTCCACAT ATGACCCATA GCCACTGGAT TAAAAAAGAT AGTTTGTCTG	60
AAGCTGAGAG AGCGGCAcCC AGGCTTATGC TAAAGAGAAA GGTTCGACCC CTCCTTCGAC	120
AGACCATCAG GATTCAAGAA ATACTGAGGC AAAAGGAGCA GAAGCTATCT ACAACCGCGT	180
GAAAGCAGCT AAGAAGGTGC CACTTGATCG TATGCCTTAC AATCTTCAAT ATACTGTAGA	240
AGTCAAAAAC GGTAGTTTAA TCATACCTCA TTATGACCAT TACCATAACA TCAAATTGA	300
GTGGTTTGAC GAAGGCCTTT ATGAGGCACC TAAGGGGTAT ACTCTTGAGG ATCTTTTGGC	360
GACTGTCAAG TACTATGTCG AACATCCAAA CGAACGTCCG CATTCAAGATA ATGGTTTGG	420
TAACGCTAGC GACCATGTTT AAAGAAACAA AAATGGTCAA GCTGATACCA ATCAAACGGA	480
AAAACCAAGC GAGGAGAAAC CTCAGACAGA AAAACCTGAG GAAGAAACCC CTCGAGAAGA	540
GAAACCGCAA AGCGAGAAAC CAGAGTCTCC AAAACCAACA GAGGAACCAG AAGAATCACC	600
AGAGGAATCA GAAGAACCTC AGGTCGAGAC TGAAAAGGTT GAAGAAAAAC TGAGAGAGGC	660
TGAAGATTTA CTTGGAAAAA TCCAGGATCC AATTATCAAG TCCAATGCCA AAGAGACTCT	720
CACAGGATTA AAAAATAATT TACTATTGG CACCCAGGAC AACAATACTA TTATGGCAGA	780
AGCTGAAAAA CTATTGGCTT TATTAAAGGA GAGTAAGTAA AGGTAGCAGC ATTTTCTAAC	840
TCCTAAAAAC AGGATAGGAG AACCGGAAAA CGAAAAATGA GAGCAGAATG TGAGTTCTAG	900
TTCTCATTTT TTTTCATGAA ATGTGCAAAA TATAGTAGAT TGAAACTAGA ATAGTATACC	960
TCTACTTCTA AAACATTGTT AGAAATCGAT TTGACTGTCC TGTTCCTTATT TCATTTTACT	1020
ATATCTTAAC AGATAGTGTA AATAAGATA AACTATTTAC TGGCTAATTA ATCAGTTAAA	1080
CACTAGTTAA GGAGTAATGA TGAAAAAAG AACAATACTA TTATTGATGG CCAGTCTGTT	1140
AGCTCTTGTC TTAGGAGCAT GTGGTTTCTT GGACATATTG ATCCTGGATC ATTCTCATCA	1200
GGATTACTCT TTA CTGCTAT TTTAGAACT GGGGTGGTTT GATGGAAAGT ATTGGTCTTG	1260
TTATCGTTTC ACATTCCAAA CACATTGCAG AAGGTGTTGT TGAAGTGAAT AGTAAAGTAG	1320
CTAAAGATGT TCCGATTACT TATGTAAGAG GAACCGAGGG CGGAGGAATT GGAACGAGTT	1380

1285

TTGAACAAGT AGATAGGGTT GTTTCGAAA ATCCAGCAGA TACTTTACTT GCCTTTTTTG	1440
ACCTAGGTTC TGCTAAAATG AACTTAAAAA TGGTGAAGTGA TTTCAGTGAT AAAAGTATCA	1500
TCATCAACAG GGTTCCAATT GTAGAAGGTG CCTATAATGC AGCTGCTCTT CTTCAGGCTG	1560
GTGCAGAACT GTCAGTTATT CAAACACAGT TaGCGGAgCt TGAAATCAAT AAATAAGGAA	1620
TTTTACTATA ACTCTTTTTA TAGATAAGCT ATTGATTATC TCAACTATAA TAATGTTAAG	1680
TrAA	1684

(2) INFORMATION FOR SEQ ID NO: 259:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 970 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

AGGAGTGGAG AnATATGAAG ACACAAATTT TCACATTATT GAAAAATCGTT GCTGAGATTA	60
TTATTATTTT GCCATTCTTA ACTAATCTAT AAGTTCTTTA TATTGCTGAA AACGCAATTC	120
AAAAAGGGCT ATTAATTGTG GATTTTCTAA TACCTGCAGA GATTGGATAA AGCGTTCAAT	180
CTCTTTTGA TTGCTTCCCT TTGTTGAAG AAAGACACTC ATCTTCTTTA AAAATGCCCA	240
CGATACTTTT TCAAAAACAT CATACGGTCG TAACATCCTC TCCAACTCGG CTTCGAAGAT	300
TGGGATGTAG GAGAAAAGTT TTCGCTCCAT GAGTTCTGAT AAGATATTTA AGAGTCCTTG	360
CTTCATATAC AATCGATTGT GTACTAACTC TTAAATTTCT TTGGATTTT CGAGTAAGGA	420
GGTTGATAAA AAAATCAGAT CTTGATTGCT CAAGAAGGGC ATGGTATTGC AAAAGAGATA	480
GAGTTCAAAC CAGGTCCAAG ACTCGATAGC ATAGAGATAG GTGGTCAAAA ACTCGCTATC	540
CTCCTCTGCT AGTGGGTAGC TTTTATTTAG TGAATGGATG GCATCTTTAA TCACGATGGC	600
ATTCAAACGA CGATAGGTCT GCGCCATCTG TTCTTGATCG ACTTCCTCCA ATAGCTGCTC	660
TAAAGCAGCT ATATCCTGAT GGGCAAAGCG ATTCACAACC TTTGACCGA TTCGATATG	720
TGGAGATTCT TGATAGTTGT TGAGCTTGTG CCCAACTCA TCAAAGGTCA CATTTATACC	780
TTGGATAGCT AGAATCAACT TATCCGAGA CAGCATAGAC TGCCCTAGTT CAACTTGGA	840
CAACTGAGAA GCTGTTAGAC CCTCACAAGC CACATCTGAC TGCTTGAGCT TTCTGCCAA	900
ACGTAATTCC TTGTAAAATT CCCCAGTTC CATTTCTCA ATCATCTGAC CACCTCCTAG	960
CTTTTGCAGG	970

1286

## (2) INFORMATION FOR SEQ ID NO: 260:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2996 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

GTTGACCACG GGTAAACTA CCCTAACTGC AGCTATCACA ACTGTTTGG CACGTCGCTT	60
GCCTTCATCA GTTAACCAAC CTAAGACTA TGCCTCTATC GATGCTGCTC CAGAAGAACG	120
CGAACGCGGT ATCACTATCA ACACTGCGCA CGTTGAGTAC GAAACTGAAA AACGTCACCTA	180
CGCTCACATC GACGCTCCAG GACACGCGGA CTACGTTAAA AACATGATCA CTGGTGCTGC	240
TCAAATGGAC GGAGCTATCC TTGTAGTAGC TTCAACTGAC GGACCAATGC CACAACTCG	300
TGAGCACATC CTTCTTTCAC GTCAGGTTGG TGTTAAACAC CTTATCGTCT TCATGAACAA	360
AGTTGACTTG GTTGACGACG AAGAATTGCT TGAATTGGTT GAAATGAAA TCCGTGACCT	420
ATTGTCAGAA TACGACTTCC CAGGTGACGA TCTTCCAGTT ATCCAAGGTT CAGCACTTAA	480
AGCTCTTGAA GGTGACTCTA AATACGAAGA CATCGTTATG GAATTGATGA ACACAGTTGA	540
TGAGTATATC CCAGAACCAG AACGTGACAC TGACAAACCA TTGCTTCTTC CAGTCGAGGA	600
CGTATTCTCA ATCACTGGAC GTGGTACAGT TGCTTCAGGA CGTATCGACC GTGGTATCGT	660
TAAAGTCAAC GACGAAATCG AAATCGTTGG TATCAAAGAA GAAACTCAAA AAGCAGTTGT	720
TACTGGTGTT GAAATGTTCC GTAAACAAC TGAACGAAGT CTTGCTGGAG ATAACGTAGG	780
TGTCCTTCTT CGTGGTGTTT AACGTGATGA AATCGAACGT GGACAAGTTA TCGCTAAACC	840
AGGTTCAATC AACCCACACA CTAAATTCAA AGGTGAAGTC TACATCCTTA CTAAAGAAGA	900
AGGTGGACGT CACACTCCAT TCTTCAACAA CTACCGTCCA CAATTCTACT TCCGTACTAC	960
TGACGTTACA GGTTCATCG AACTTCCAGC AGGTACTGAA ATGGTAATGC CTGGTGATAA	1020
CGTGACAATC GACGTTGAGT TGATTCACCC AATCGCCGTA GAACAAGGTA CTACATTCTC	1080
TATCCGTGAG GGTGGACGTA CTGTTGGTTC AGGTATGGTT ACAGAAATCG AAGCTTAATT	1140
CGATTTAGTT CCCAGAAGAA CAATTATTTA AGTTAGACAC TAAAAGAATC TTGCTTGGCA	1200
AGGTTCTTTT TTTAGATATT GAACTAATAC TCAATGAAAA TCAAAGAGCA AACTATAATA	1260
TATTGAAACT AGAATAGTAC ACATCTACTT CTAAACATT GTTAGAAATC GATTTGACTG	1320
TCCTGATCGA TTTGTCTTGT TCTTATTTCA TTTTACTATA GAAAGTTAGC TACAGACTGC	1380
TCAAAACATT GTTTTATAGT TGTAGATAGA ACTGACGAAG TCAGTAACAT CTATACGACA	1440

1287

AGGCCAAGCT GACCGGTTT GAAGAGATT TCGAAGAGTA TAATACTAGA CTAAATCAA	1500
AAAGCATTAT ACAATAGTAA TATGAAATCA ATTAAAGAAG AAATCCAAAC CATCAAAACA	1560
CTTTTAAAG ACTCTCGTAC AGCTAAATAT CATAAACGCC TTCAAATCGT TCTATTTTCGT	1620
CTGATGGGCA AATCTTATAA AGAGATTATA GAACTTTTAT AGTGGTTTGA AATAAGATGT	1680
GAACAACTCT ATCAGGAAAG TCAAACTAAT TTATAGAAAT ATTTTAGCAG CCAAGGTGTA	1740
CTGTTATAGA TTCAATACAC TTTAGACTGT AATCAAAACA CGATTGCGG AAATGTAAAA	1800
AATATGAGGA GTTCGGACTC GACTCTCTCC TTCAAGAAAC ACGTGGTGGT CGTAACCATG	1860
CTTATATGAC GGTGAGCAA GAGAAAGTCT TTCTTGCCCG CCATTGAAG GCTACAGAGG	1920
CAGGAGAATT TGTACAATT GATGCCTTAT TTCAGGCTTA TAAAAAGGAG TTAGGTCGTT	1980
CCTACACACG TGATGCCTTC TATCAACTGT TGAAGCGCCA TGGTTGGCGA AATATTACGC	2040
CACGTCCAGA ACATCCTAAG AAAGCAGATG CTCAAACCAT TGTCGCGTCT AAAAATAAAG	2100
TCTCAATTCA AGAAGACAAG TGAAGTGCAC CCAAAAGTT AGACAGAAAA AATCTAACTT	2160
TTGGGGTGTT TTTATTATGA AATTAACTTA TGATGATAAA GTTCAGATCT ATGAACTTAG	2220
AAAACAAGGA TATAGCTTAG AGAAGCTTTC AAATAAATTT GGGATAAACA ATTCTAATCT	2280
TAGGTACATG ATTAATGA TTGATCGTTA CGGAATAGAG TTCGTCAAAA AAGGAAAAAA	2340
TCGTTACTAT TCTCCTGATT TAAAACAAGA AATGATTCAT AAAGTCTGAC ATGAAGGCTG	2400
GACTAAAGAT AGAGTTTCTC TTGAATACTG TCTCCCAAGT CGTACGATAC TTCTTAACTG	2460
GCTAGCACAA TACAGGAAAA ACGGGTATAC TATTGTTGAG AAAACAAGAG GGAGAGTACC	2520
TGAGAGCGGA GAATGCCATC CTAAAAAGT TAAGAGAACT CCGATTGAAG GAGGAAAAAG	2580
AGAAAGAAGA AAGACAGAAA TTATTCAAGA ATTAATGACT GAGTTTTTCGT TAGATATTCT	2640
TCTAAAAGCC ATTAACTAG CTCGTTTGAC CTACTACTAT CACTTGAAAC AGCTAGATAA	2700
ACCAGATAAG GACCAAGAGC TTAAAGCTGA AATCAATCC ATTTTATCG AACACAAGGG	2760
AAATTATGCT TATCGTCGGA TTTATTAGA ACTAAGAAAT CGTGGTTATC TGGTAAATCA	2820
TAAAAGAGTT CAAGGCTTGA TAAAAGTACT CAATTTACAA GCTAAATGC GACAGAAACG	2880
AAAATATTCT TCTCATAAAG GAGACGTTGG CAAGAAGGCA GAGAATCTCA TTCAAGGACA	2940
ATTTGAAGGC TCTAAAACAA TGGAAAAGTG CTACACAGAT GTGACAGAAT TTGCCG	2996

(2) INFORMATION FOR SEQ ID NO: 261:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 837 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double

1288

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

CTTATCAACT CCGACATGG CTCTCAGACC AATCCAAATC CCTAAAAAA TCAGAACAAG	60
GATGGTGGTC AAGATCAAAC TCTCGAAATA TAAAGAAAAT AGTTGCAGTA GCATGATTTC	120
TCTCATTTCT ATCTTTTSTA AAGAGTAAAC TCAGCTAGTC CAACTAACTG AGTTTTCCTT	180
TATCTATTAT ATCAAATATA AGTCCGTTG TAACTAGCGA AGAATTCCTT TGCCCGCTCT	240
TCTTTAGGGG TGTGGATAAT CTCATCCGGA GTTCCAGACT CGATGATTTT CCCCTTATCT	300
AAGAAGAGAA TTTTATCCGC AACTTGGGCT ACAAAGGACA TGTCACTACT GACCAAAATC	360
ATGGTCTGAC CTGACTTAGC AGCATCTGCA ATAGACTTTT CTACTTCACC GACCAATTCT	420
GGGTCAAGGG CTGAAGTTGG TTCGTCTAAG AGCAAAACAT CTGGTTTCAT AGCAAGCGCA	480
CGCGCTAGGG CAACCCGTTG CTTCTGTCCA CCTGATAAAT GCGGAGGATA ATGGTTTCA	540
CGGTCCGAAA GCCCAACCTT AGCCAACTCT TCCTTGGCAA TCTTAGTCGC TTCTTGGTCA	600
GATAATTCTT TGACAACAAC CAAGCCTTCT TTCACATTAT CAAGTGCTGT TCGGCGTTCA	660
AACAAATTAA ACTGTTGGAA AACCATAGAC AACTTACGAC GTAGGGCAAG GATTTCTTCT	720
TGAGTGATTT TAGAAAAATC AACTGAAAAA CCATCAATCT GAATAGAGCC ACTGTCAGGT	780
GTTTCTAGAT AATTGAGACT GCGAGAAAGG TTGATTTTCA GCTCTGAAGA CCAATCA	837

(2) INFORMATION FOR SEQ ID NO: 262:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 868 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

CCGAACAAAA TGGGCTAATT AGATTATAGT AAGAAAGGTA AGTTAAAAAT GAGAATTGCA	60
ATTGGATGTG ACCACATCGT AACTGATGAA AAAATGGCGG TTTCAGAATT TTTGAAATCA	120
AAAGGATATG AAGTCATTGA CTTTGGTACC TATGACCATA CACGGACTCA CTACCCAATC	180
TTTGGTAAAA AAGTAGGGA AGCTGTAAC AGCGGTCAAG CTGATCTTGG AGTATGTATC	240
TGTGGTACTG GTGTTGGTAT CAACAACGCT GTAAATAAAG TTCCAGGTGT TCGTCTGCC	300
TTGGTTCTGT ATATGACAAC AGCCCTTTAT GCTAAAGAAC AATTGAACGC TAACGTTATT	360
GGTTTTGGTG GTAAAATTAC TGGTGAATTG CTTATGTGTG ATATCATCGA AGCTTTCATC	420



1289

CATGCTGAAT ACAAACCAAC TGAAGAAAAC AAAAAATTGA TTGCGAAAAT TGAACATGTT	480
GAAAGTCACA ATGCTCAACA AACAGACGCA AACTTCTTTA CAGAATTCCT TGAGAAATGG	540
GATCGTGGAG AATACCACGA CTAAGAGGTG ACCTATGATT TTAACAGTCA CAATGAACCC	600
ATCCATCGAT ATTTCCTATC CCTTGGATGA GTTGAAGATT GATACTGTCA ATCGTGTGGT	660
GGATGTAACC AAAACGGCTG GTGGTAAGGG ACTCAATGTT ACCCGAGTAC TTTCAGAATT	720
TGGCGATTCT GTTCTTGCTA CTGGTTTAGT GGGTGGCAAA CTGGTGAGT TTTTGGTTGA	780
ACATATCGAT AATCAAGTAA AGAAAGATTT CTTCTCAATT AAGGGAGAAA CTCGTAAC TG	840
TATCGCTATT CTCCACGGAG ACAACCAA	968

(2) INFORMATION FOR SEQ ID NO: 263:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3744 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CCGTTCAAAG TCTTCATAAG ACTCGAAAGT CACAGTTCTT TCGTTCTTGC TGGCATCTAT	60
ATAGGTAATT TCAATCATGT TTAAACTCC TTTGTTTAAT GCTAACTTTA TTTTACTCCT	120
TATAAAGAG AATGTCAAGA AAAATGATTG CGCACGCAAC TTTTTTTAAA ATCATCTTAA	180
ATCAAGAAAT CCAAACCTGC TTCCAAGCTT TCTTCGACAG TCTTTTGTAG CGAGGCCAGT	240
GTCTTTTGCC CATCATTTGT CAGGCAGATA AACTAGAGC GTCTATCTTG ATGGCAACAC	300
ATGCGACTGA GTAGACCGCA ATTTT TAGCT TCCAAGCGAG CCACCATCCT AGAAACTGCG	360
CTCGGGCTCA GATGAAGCTT ATCTGGCAGG TCAATCTGGC GTAGAGATTT TTCTTCAGCC	420
AAGTCCAGAT AGTAGAGCAG GTAGAACTCT TTCAAGGTCA GACTTTGCTC GCTCTGTTGG	480
GCAATGGTCT CTTCCAAGAG ACTTTCAATT TCTTTCTGAC GCCGATTGAA GTCAAACCAT	540
TTTTCCAAAT AGGTCATAGT GTCTCCTTTC TTTT TAGAGT CATAAATAGA AGAAAGTCCA	600
TTAACGGGCA GTCTCTGCGT CACAAGATGA TTGCGCATGC AATAATTATA CTACTTTTCA	660
AGAATGCTGG CAAGCTCTGT TTTT TAGTGG TTTTATTTT GTGTGAATAA TGGGGGAATC	720
CTATTGTTTC AATTTCTAAC TCCTTATCAC ATTCGAATTC AGATTTTATT TCATTTCTCT	780
ATCTATAGTT GCTTAGTTTA AAATAAGCAT GGTCTAATAA AGCTATGCAT ATAGTACTGA	840
TTTTAAACAA GGAGCATTAG ATTCCATTAA AGGAGGGCAC AGACATGTCG AGGCGGCCAA	900

1290

AGTTTTTGAT GTCGGCGTCA GAACTCTCTT CACGTGGGAA AAGAAAGACG TAAACAAGGG	960
AACTTAGAGC GGAAAAAGCG AGTCGTCAAA AAGCGTAAGA TCCCTTTAGA AGAATTGAAA	1020
GCCTTTGTAG AGGCTCATCC AGACGCTTTT TTACGGGAAA TTGCGGCCCG TTTTGATTGT	1080
GCTTTGCCCT CCGTATGGGC AGTTTAAAG CAGATTAAGG TCATTTTAAA AAAGACGACC	1140
AGTTTTAGGG AACAAAAGCC TGAGAAAGTT TCTGAGTTTC TTGATATTTT GGATAACCTA	1200
AAAGATTTAC CAGTCCTATA TATTGACGAA ACGGGAATCG ACCGCTACCT CTATCGTCTT	1260
TATGCAGGGG CTCCTAGAGG GGAGAAAGTC TATGGCAAGA TTAGCGGACG GCGTTTTGAG	1320
CGGACTAATG AGGTGGAGCA AAAACTCAAT GGTAAGTTTC TAATCAGATA TATTGATTCA	1380
CAAATTAGAG AATGAAAGAA TAATTATGCA TAAAAATAGG AATATAAACC AAAAATTAGC	1440
TGATTTATAC TCATTTGCGT GTCTTTATAA AAAACTTATC TTATAATATA TATATATATA	1500
TATACAAAAT AGTAAATGC TTTTTTTTTT TAGCAAAAAT ACCTCAAGTT TCTTGCTATT	1560
TTGGGTTCCT TATTCTATAA TTATAGTATG GTAATTTATT TATATCCATA CATGAAAATA	1620
ATACTCGAAA GGAAATTTCA AAATATTTT TAGACGTCAG AAGGGTGAAT ATAGAGAAAC	1680
AGACCGAGTA ACTCGGTTCA AATTAATCAA ATCAGGGAAG CATTGCTAC GGGCCTCGAC	1740
TTCTCTTTTT GGCTTGTTTA AGGTCTTGGC AGGTGGTGT GATACTACTC AGGTCATGAC	1800
CGAAACGGTA GAAGATAAAG TAAGTCATTC AATTACTGGG CTTGATATCC TCAAGGGGAT	1860
AGTTGCTGCG GGAGCTGTCA TAAGTGAAC CGTTGCAACT CAAACGAAGG TATTACAAA	1920
TGAGTCAGCA GTACTTGAAA AACTGTAGA GAAACGGAT GCTTTGGCAA CAAATGATAC	1980
AGTAGTTCTA GGTACGATAT CTACAAGTAA TTCAGCGAGT TCAACTAGTT TGTGAGTTT	2040
AGAGTCGGCA AGTACATCTG CATCTGAGTC AGCCTCAACC AGCGCTTCGA CCTCAGCAAG	2100
TACAAGTGCA TCAGAATCAG CAAGTACATC GGCTTCGACA AGTATTTCTG CATCATCTAC	2160
TGTGGTAGGT TCACAAACAG CTGCCGCTAC AGAAGCAACT GCTAAGAAGG TCGAAGAAGA	2220
TCGTAAGAAA CCAGCTAGTG ATTATGTAGC ATCAGTTACA AATGTCAATC TCCAATCTTA	2280
TGCTAAGCGA CGCAAGCGTT CAGTGGATTC CATCGAGCAA TTGCTGGCTT CTATAAAAAA	2340
TGCTGCTGTT TTTTCTGGCA ATACGATTGT AAATGGCGCC CCTGCAATTA ATGCAAGTCT	2400
AAACATTGCT AAAAGTGAGA CAAAAGTTTA TACAGGTGAA GGTGTAGATT CGGTATATCG	2460
TGTTCCAATT TACTATAAAT TGAAAGTGAC AAATGATGGT TCAAAATTGA CCTTTACCTA	2520
TACGGTTACG TATGTGAATC CTAACAACAA TGATCTTGGT AATATATCAA GTATGCGTCC	2580
TGGATATTCT ATCTATAATT CAGGTACTTC AACACAAACA ATGTTAACCC TTGGCAGTGA	2640
TCTTGGTAAA CCTTCAGGTG TAAAGAACTA CATTACTGAC AAAAATGGTA GACAGGTTCT	2700

1291

ATCCTATAAT ACATCTACAA TGACGACGCA GGGTAGTGGG TATACTTGGG GAAATGGTGC	2760
CCAAATGAAT GGTTCCTTTG CTAAGAAAGG ATATGGATTA ACATCATCTT GGA CTGTACC	2820
AATTACTGGA ACGGATACAT CCTTTACATT TACCCCTTAC GCTGCTAGAA CAGATAGAAT	2880
TGGAATTAAC TACTTCAATG GTGGAGGAAA GGTAGTTGAA TCTAGCACGA CCAGTCAGTC	2940
ACTTTCACAG TCTAAGTCAC TCTCAGTAAG TGCTAGTCAA AGCGCCTCAG CTTCAGCATC	3000
AACAAGTGGC TCGGCTTCAG CATCAACCAG TGCCTCGGCT TCAGCGTCAA CCAGTGGCTC	3060
AGCTTCAGCA AGTACCAAGT CTTCAGTCTC AGCATCAACA AGTGCTTCAG CCTCAGCATC	3120
GACAAGTGCC TCGGCTTCAG CAAGCACATC AGCATCTGAA TCAGCGTCAA CCAGTGCTTC	3180
GGCTTCAGCA AGTACCAAGT CTTCAGCTTC AGCATCAACC AGCGCCTCGG CCTCAGCAAG	3240
CACCTCAGCT TCTGAATCGG CCTCAACCAG CGCCTCGGCC TCAGCAAGCA CCTCAGCTTC	3300
TGAATCGGCC TCAACCAGCG CCTCAGCCTC AGCATCAACG AGTGCTTCGG CTTCAGCAAG	3360
CACAAGCGCC TCGGCTTCAG CATCAACGAG TACGTCAGCT TCAGCGTCAA CCAGTGCTTC	3420
AGCCTCAGCA TCAACAAGTG CGTCAGCTCA GCAAGTATCT CAGCGTCTGA ATCGGCATCA	3480
ACGAGTGCGT CTGAGTCAGC ATCAACGAGT ACCTCAGCCT CAGCAAGCAC CTCAGCTTCT	3540
GAATCGGCCT CAACCAGTGC GTCACCTCAG CATCGACAAG CGCCTCAGCT TCAGCAAGTA	3600
CCAGTGCTTC AGCCTCAGCG TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG	3660
CATCAACCAG TCGCTCAGCC TCAGCAAGTA CTAGTGCATC GGCTTCAGCA TCAACCAGTG	3720
CCTCGGCTTC AGCGTCAAAC AGTG	3744

## (2) INFORMATION FOR SEQ ID NO: 264:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 795 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

CGATAAAGAG GCCTTGAGTA ATCTCAATTT GCAGATTGAA AATGGAGAGA TTATGGGCTT	60
GATTGGTCAT AATGGGGCTG GAAATCGAC CACTATAAAA TCCCTAGTCA GTATCATTTT	120
ACCCAGCAGT GGTGCTATTT TGCTAGACGG TCAGGAGTTA TCGGAAAATC GCTTGGCTAT	180
TAAACGAAAG ATTGGCTACG TAGCAGACTC GCCTGACTTA TTTTACGCT TAACGGCCAA	240
TGAATTTTGG GAATTGATCG CCTCATCCTA TGATCTGAGT AGATCTGACT TGGAGGCTAG	300

1292

TCTAGCTAGG CTATTGAACG TTTTGTGATT TGCTGAAAAT CGCTATCAGG TTATTGAAAC	360
TCTTTCTCAC GGAATGCGTC AGAAAGTCTT TGTCATCGGA GCACTCTTGT CTGATCCCGA	420
TATTTGGGTC TTGGATGAAC CCTTGACTGG TTTGGATCCC CAGGCTGCCT TTGATTTGAA	480
ACAGATGATG AAGGAACATG CACAAAAAGG GAAGACAGTC TTGTTTTCAA CTCATGTCCT	540
AGAGGTGGCA GAGCAAGTCT GTGATCGGAT TGCCATTTTG AAAAAGGGGC ATTTGATTTA	600
TTGTGGTAGT GTAGAGGACT TGAGAAAAGA TTACCCAGAC CAGTCTTTGG AAAGTATCTA	660
CCTTAGTCTT GCTGGTAGAA AAGAGGAGGT TGCGGATGCG TCTCAAGGTC ATTAAAAAAT	720
TAGTTGATAT CAATATCCTT TATTCATCTC AAGAAGCTAA TCTGGCTAAT CTACGAAAGA	780
AGCAGGCTAA GAATC	795

(2) INFORMATION FOR SEQ ID NO: 265:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2231 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

TGTAATGTG CTTGGCAGCw TCCTTGACAC TGCTACTACC ATTTCCCATTA GCGACCGACA	60
TACCAACGCC AGCCAGCATT TCAAGATCAT TATCTGAGTC ACCAAAAGCC ATGACTTGGT	120
TGAGGTCAAA GCCATATTCT TTCCCAACTC GGCGAATGCC TTCTAATTTA GAATTTCCCT	180
GATTGATGAC ATCCGATGCA AAAGGATTGC TACGTGTCAA TTTCAAGTCT TCAAAATCAG	240
CTGCCGCCTT CTCAGATTCT TCTGGTGTCA TCAGCATCAA AACTTGGTAG ATAGGCTGAT	300
TCATCAGGTG AAGCAGGTCC TCTTCCTTTT GGGGAACAAC CTTGCTGACC ATGCGATTAA	360
AAGACTGACT CACCGTCCGA GTTAAAACAG AGGGAACGAA GCGACTAATT CGTTGGGAAA	420
AAGAACCAG ACCAAAGGAC ATGATTTTAG AACCCAACAT GGCATCCTTG GTCCCTAGAG	480
CAATCTCCGT GCCCTCTTTT TTAGCATAGC TAATTAGATG GCGCAATGT AACTTGGAAA	540
TAGGGCTCGT GAACAAGACT CTGTCTTTAC TAAAGATATA CTGGCCATTA TAGGTTACCG	600
CAAAATCCAG ATCCAAATCG TCCATCAATT CCTTAACAAA AAAAGGTCCT CGCCCTGTCTG	660
CTACGCCAAC TAGTACCCCT TGTCTTTGA CAATCTTAAT CGCATCCTTA GTGGATTTC	720
AAACACTCTT GCGATTGTTG ACCAAGGTTT CATCGATATC AAAAAAACA GCTTTGACTT	780
CCATCCTATC CCAATCTCCC CTTTGTGAT ACAATGATTA TACCACATT CAGAAAGAGT	840
GAGTAAATCA TGCCTAAGAA AATCCTTGTT TTACATACGG GTGGAACTAT TTCCATGCAG	900

1293

GCCGATGCTT CTGGCGCTGT TGTGACGAGT TCAGATAATC CCATGAACCA TGTGTCCAAC	960
CCACTTGAAG GAATCCAAGT CCACGCCTTG GACTTTTTTA ACCTTCCAAG TCCCCATATC	1020
AAACCCAAAC ATATGCTGGT CCTCTACCAG AAAATTAAAG AGGAAGCAGA TAACTACGAT	1080
GGAGTGGTGA TCACACACGG AACCGATACT TTAGAGGAAA CAGCCTATTT CTTGATACC	1140
ATGGAAGTTC CCCATATGCC TATCGTTCTA ACAGGAGCCA TGGGTACCC AATGAGCTCG	1200
GTAGTGATGG TGTTTATAAT TACCTAAGTG CTTTACGAGT GGCCAGCGAT GACAGGGCTG	1260
CTGACAAAGG AGTTTGGTC GTTATGAACG ATGAAATCCA CGCTGCCAAG TATGTCACCA	1320
AAACACATAC GACTAATGTC AGCACCTTCC AGACTCCAAC ACATGGCCCC CTTGGTCTCA	1380
TCATGAAACA GGAAATCCTC TACTTCAAAA CAGTGAAACC TCGTGTTCGC TTTGACCTTG	1440
ATCACATACA AGGTTTAGTC CCTATCATCT CGGCTTATGC TGGTATGACA GATGAGCTGA	1500
TTGATATGCT GGATTTAGAA CACTTGGACG GTTTGATTAT CCAAGCCTTC GGAGCTGGTA	1560
ATATTCCCAA AGAAACGGCT CAAAAATTAG AAAGCCTTCT GCAAAAAGGA ATTCCAGTCG	1620
CTCTGGTATC ACGATGCTTT AACGGTATTG CCGAGCCTGT TTATGCATAC CAGGGTGGGG	1680
GCGTACAGTT GCAAAAAGCA GCGTTTTCT TTGTTAAAGA ACTCAACGCC CAAAAAGCTC	1740
GCTTGAAACT CCTCATCGCC CTCATGCCG GACTAACAGG ACAGGCTTTG AAAGACTATA	1800
TGGAAGGCTA ATACTCTTCG AAAATCTCTG CAAACCAGT CACGTCGCCT TACCGTATGT	1860
ATGGTACTGA CTTGTCAGT TTCATCTACA ACCTCAAAAA CATGTTTTGA GCTGACTTCG	1920
TCAGTTCTAT CTACAACCTC AAAACATGT TTTGAGCTGA CTTGTCAGT TCTATCTACA	1980
ACCTCAAAAA CATGTTTTGA GCTGACTTCG TCAGTTCTAT CTACAACCTC AAAACATGT	2040
TTTGAGCTGA CTTGTCAGT TCTATCTACA ACCTCAAAAA CATGTTTTGA GCTGACTTCG	2100
TCAGTTCTAT CTACAACCTC AAAACATGT TTTGAGCTGA CTTGTCAGT TCTATCTACA	2160
ACCTCAAAAA CATGTTTTGA GCTGACTTCG TTAGTTTCAT CTACAACCTC AAAACATGT	2220
TTTGAGCTGA C	2231

(2) INFORMATION FOR SEQ ID NO: 266:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1310 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

1294

GAGTCAAAGG CTCCGAGGTT GACTTTTTTAC AAGGGGACAG GTGAATATTA TCTAGACCTG	60
TCAGAAATTC TCTTCTTTGA AACAGAAGGG AGCAAGATCT ACGCTCATAA CCAGAAGGAA	120
GCTTATGAGG TTCGCCTCAA GCTCTATGAG TTGGAGTCTA TCTTGCCTCG CTATTTTAAT	180
CGAGTTTCCA AGTCAACGAT CGCAAACATC CGTCAGATTT ACTCAGTGGG CAAGTCCTTT	240
TCAGGAACGG GCACCATTTT CTTTATCAG ACGCACAAGG AGGTTTCATGT CTCACGGCAT	300
TACCAATCCC TCCTAAAAGA AAATCTAAGA AACATGAGGT AAAAAACATG AAAAGAAAG	360
CATTTGGTAT TGTTTTATTG GTTTTAGCAG CTTGGATCTT GCTGCAAGGG AATTTTGGAA	420
TTCTTCTTTT GGATGGTAAA ATATGGCCTT TACTAGGTAT TGTTTTTTTT GCTTATAAGT	480
CCATTGAGTC CATCCTTAGA CGTCATCTCA CTTGGGCAGT TTTACAGGT TTAGTGGCGC	540
TCATCATTGC AAATTACGCT TATGACTTGT TACCAGTTAC CAATCATTCT CTTATTGGG	600
CTAGCATCTT GGTGGTACTT GGTGTTGGTT ATCTGACGCA TTCAAGTAAG TTCTGGAATG	660
AAAAAAAATG GTGGTACAAT GGGAAAAAAA CAGTCGTCAC GGATAAGGAA GTCGCTTTTG	720
GTAGCGGGAC CTTCTATAAG CAAGATCAAG ATCTCGTAGA TGACCAAGTG GAAGTCGCTT	780
TTGGGGATGC TAAAATCTAC TATGATAATG CAGAGATGCT AGGTGATTTT GCAACTTTAA	840
ATATTGAAGT GGCCTTCGGG AATGCAACCG TCTATGTTCC ACAACACTGG CGTGTAGATT	900
TGAAAGTAGA AACCTCCTTT GGTGCAGCTA AGGCTGACGC TCCTGTAGCC CCAACCAGCA	960
AAACCTTGAT TATCCGTGGA GATGTGGCTT TTGGGAAGTT GGAAATTGTC TACGTTAAAT	1020
AAAAAATCT TCACTTCAAC CATCAAAATA GACGTACTAA GAGTAGGAAA TTGATGCCTT	1080
GCTCTGATTT CAGTCTATG GTTGTTAGAC TTAAAAAAT GAAATGCTGC CTTTAAAGT	1140
TGTATATTTT TCGATATTTT GGCTTTTACG TTTGATGTAT CTATGTACTA CAGCGTAGAT	1200
GATGTAGTGT CAAATGCTTT TAAAAACGG ATGATATTGG ACAGTTTTTT TGCCTTTAAT	1260
TGCTCAGGAA CCATGAAAGT CAGTACCTGG GTTTATGACA AGGGAGAATG	1310

## (2) INFORMATION FOR SEQ ID NO: 267:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5922 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

ACTCTGATTT GATTGGAACG ACAGTCGGTG CCATTGCAGT TACTTCAAAC GTAACGACTT	60
ATGTTGAGTC TGCTGCTGGT ATCGGTGCAG GTGGACGTAC TGGTTTGACA GCCTTGCTTG	120

1295

TAGCTATCTG TTTTGGGATT TCAAGCTTCT TTAGCCCACT TCTAGCGATC GTACCAACAG	180
CGGCTACAGC TCCAATCTTG ATTATCGTTG GGATTATGAT GCTTGGTAGC TTGAAAAATA	240
TCCATTGGGA TGATATGTCT GAAGCAGTTC CTGCCTTCTT CACATCTATC TTTATGGGAT	300
TCAGCTACTC TATCACTCAA GGGATTGCAG TTGGTTTCTT GACTTACACT TTGACTAAGC	360
TTGTAAAGG TCAAGTTAAA GATGTTTCATG TCATGATTG GATTTTGGAT GCCTTGTTTA	420
TCCTTAACTA CATCAGCATG GCCTTATAAT AGAATGACCC AGGGGGATT CCCCCCTTTT	480
TTAATACAAG GAGATAGGTG ATGAAAGAGA AAAATATGTG GAAAGAATTG TTGAATCGTG	540
CAGGCTGGAT TTTGGTCTTT TTAAGTCCCG TCCTTTTATA TCAGGTTCCC CTAGTGGTTA	600
CCTCTATTTT GACTTTAAAA GAAGTAGCCC TGCTACAGTC AGGGCTGATA GTTGCTGGCC	660
TTTCAATTGT GGTTCGGCT CTATTTATTA TGGGAGCTCG TAAAACCAAG TTAGCTAGTT	720
TTAATTTTC TTTTTTAGA GCTAAAGATT TGGCAGGTT GGGCTTGACT TATCTAGTTA	780
TTGTGGGTC AAATATACTT GGTTCATT TATTGCAACT GTCAAATGAG ACGACAACAG	840
CTAACCAGTC TCAGATTAAT GATATGCTTC AAAATAGTTC GTTGATTCC AGTTTCTTCT	900
TGCTAGCCTT GCTTGCTCCG ATTTGTGAGG AAATCTTGTG TCGTGGGATT GTTCCTAAAA	960
AGATTTTCCG AGGCAAGGAG AACTTGGGAT TTGTAGTCGG TACGATTGTG TTTGCTTTAT	1020
TGCATCAACC AAGTAATTTA CTTCTTTTAT TGATTTATGG AGGTATGTGC ACAGTTCTAT	1080
CTTGGACAGC CTACAAGACC CAACGTTTGG AAATGTCGAT CTTGCTTCAC ATGATTGTTA	1140
ATGGGATTGC TTTCTGTTG TTGGCTCTTG TGGTGATTAT GAGTCGGACA TTAGGAATTT	1200
CTGTTTAAAA GTTTTATGT AGGAACCGAC CTCCTTCTAC CAGGGAAAGA TGAATGCAAT	1260
CGTGTCATC TTTTCTTTT TATGGTAAAA TAGAAAAATA ATATGATGAA AATCCTTGAG	1320
GGAGTGACCG ATATGTCAAG TAAAGCCAAT CATGCAAAGA CAGTTATTTG CGGAATTATC	1380
AATGTAACCC CAGACTCCTT TTCGGACGGT GGTCAATTTT TTGCTCTTGA GCAGGCGCTC	1440
CAGCAGGCTC GTAAATTGAT AGCAGAAGGA GCCAGTATGC TAGATATCGG CGGAGAATCC	1500
ACTCGGCCGG GAAGTAGCTA TGTGAGATA GAAGAGGAAA TCCAGCGTGT TGTCCAGTG	1560
ATCAAAGCGA TTCGCAAGGA AAGTGATGTC CTCATCTCTA TTGATACTTG GAAGAGTCAA	1620
GTAGCAGAGG CTGCTTTGGC TGCTGGTGCC GATCTAGTCA ATGATATCAC TGGTCTTATG	1680
GGTGATGAGA AAATGGCTTA TGTGGTAGCT GAAGCGAGA CGAAAGTGGT CATCATGTTT	1740
AACCCAGTTA TGGCTCGACC TCAGCATCCT AGTTGCTTA TCTTCCCTCA TTTTGGTTTT	1800
GGTCAAACCT TTACAGAAAA AGAGTTAGCT GACTTTGAAA CATTGCCAAT CGAAGACTTG	1860

1296

ATGGTGGCTT TCTTTGAACG AGCACTAGCG AGAGCGGCAG AAGCTGGTAT TGCACCAGAA	1920
AATATCCTGT TGGATCCAGG AATTGGCTTT GGTCTGACCA AGAAAGAAAA TCTGCTTCTT	1980
TTACGGGACC TGGATAAACT ACATCAGAAG GGCTATCCAA TCTTTCTCGG AGTGTGCGGC	2040
AAGCgATTG TCATCAATAT CCTAGAGGAG AATGGTTTTG AAGTCAATCC TGAGACAGAG	2100
CTTGTTTTCC GAAATCGGGA CACGGCTTCG GCTCATGTAA CTACTATCGC TGCAGACAG	2160
GGTGTAGAAG TGGTGC GCGT GCATGACGTA GCTAGTCACA GGATGGCAGT TGAAATTGCC	2220
TCTGCCATTC GTCTGGCTGA TGAAGCGGAA AATTTAGATT TAAAACAATA TAAATAAGAT	2280
GAAAGAAATT GAAAACAATC AGTGGATTGC TAACTACCGG ACGGATCAAC CGCATTTTGG	2340
CTTGGAACGA ATGGTGAAC TCTTAGCTTT GCGTGGCAAT CCCCATCTCA AACTCAAGGT	2400
CCTCCATATC GGAGGGACTA ACGGCAAGGG CTCGACTATT GCTTTTTTGA AAAAGATGCT	2460
AGAAAAGCTA GGGTTGAGAG TTGGCGTGT TAGCTCGCCC TATCTCATTC ATTACACAGA	2520
CCAGATTAGC ATCAATGGGG AATCGATCTC AGAAGCGAGG CTAGAAGCTC TCATGGCAGA	2580
CTATCAGTCT TTGCTGGAGG GAGAAGCGGT CGCCAATTTA CAGGGCACAA CCGAGTTTGA	2640
GATTATCACA GCCCTGGCCT ATGACTACTT TGCCTCAGAG CAAGTAGATG TGGCCATCAT	2700
GGAAGTTGGC ATGGGTGGAC TTTTGGATAG TACCAATGTC TGTCAGCCCA TTTTGACAGG	2760
AATTACAAC ATTGGCTTGG ATCATGTGGC TCTACTTGGT GACACCTTGG AGGTCATAGC	2820
AGAGCAGAAG GCAGGTATTA TCAAACAAGG GATGCCCTTG GTAACAGGGC GTATTGCTCC	2880
AGAAGCCTTG GCTGTGATTG ACCGCATTGC GGAAGGGAAA GATGCGCCGA GACTTGCTTA	2940
CGGGACAGAT TATCAGGTTT GTCATCAAGA AAGTGTGGTG ACAGGGGAAG TCTTTGACTA	3000
TACAAGTGCT GTCAGACAAG GTCGCTTCCA GACTAGCCTG CTTGGTTTGT ACCAAATAGA	3060
GAATGCTGGG ATGGCCATAG CTTTACTTGA TACTTTTTGT CAAGAAGATG GTCGAGAGCT	3120
AGCAAGCAAT GATTTTCTTG GTCAAGCCTT GGAAGAAACA AGTTGGCCAG GGCCTTTGGA	3180
AATCGTGCA AGAGATCCCT TGATGATTTT GGATGGAGCC CACAATCCCC ATGCTATCAA	3240
GGCCTTGTTG GTAACCTTGC AAGAACGTTT TCGGATTAT CATAAGGAAA TCCTCTTCAC	3300
TTGTATCAAA ACCAAGGCCT TGGAGGATAT GTTGGACTTG CTGGGAGCCA TGCCAGTTAC	3360
CGAGCTTACT CTAACACATT TTGCGGATAG TCGGGCGACG GATGAAAACG TGCTGAAAGA	3420
GGCAGCTAAG TCTAGAAATC TCAGCTACCA AGATTGGCAT GATTTTCTAG AGCAGAATTT	3480
GACAGATAAA AAAGAAGAGA AACAAACAGT TAGGATTGTC ACAGGTTCTT TGTATTCTT	3540
GAGCCAAGTG AGGGCCTATC TGATGGAGAG GAAGAACGAG AATGGATACA CAAAAGATTG	3600
AAGCGGCTGT AAAAATGATT ATCGAGGCTG TAGGAGAGGA CGCTAATCGC GAGGGCTTGC	3660



1297

AGGAAACACC TGCTCGTGTA GCCCGTATGT ATCAAGAGAT TTTTTCAGGT CTGGTCAAAA	3720
CAGCAGAGGA ACATTTGTCA AAATCCTTTG AAATTATTGA CGATAATATG GTGGTAGAAA	3780
AGGATATCTT TTCCATACC ATGTGTGAAC ACCACTTCTT GCCATTTTAT GGTAGAGCGC	3840
ACATTGCCTA CATTCCAGAT GGTCGTGTGG CAGGCTTGTC TAAGCTAGCC CGTACGGTTG	3900
AAGTTTATTC GAAAAACCA CAAATTCAAG AACGTTTGAA TATCGAAGTG GCCGATGCCT	3960
TGATGGACTA TCTAGGTGCT AAAGGAGCCT TTGTGTGCAT TGAGGCGGAA CATATGTGTA	4020
TGAGTATGCC TGGTGTTAGA AAACCAGGCA CTGCAACCTT GACGACAGTA GCTCGTGGTC	4080
TATTTGAAAC AGATAAGGAT CTCCGTGACC AAGCTTATCG TTTAATGGGG CTATAAAAAG	4140
AATCCGCTTC AAGCGGATTT TTCTAGAAAG GAATCATTAT GGATCAACTG CAGATTAAGG	4200
ATTTGGAAAT GTTGCCTAT CATGGTCTTT TTCCTAGTGA GAAAGAATTG GGGCAGAAAT	4260
TTGTGCTTTC AGCCATCCTA TCCTATGATA TGACCAAGGC AGCTACAGAC TTGGATTTAA	4320
CAGCCTCTGT CCATTACGGA GAATTGTGTC AGCAGTGGAC GACTTGGTTT CAGGAAACGA	4380
GTGAAGATTT GATTGAAACG GTAGCCTATA AACTGGTGGA ACGTACCTTT GAGTTTTATC	4440
CTCTTGTCCTA AGAAATGAAG TTGGAACCTG AAAAACCTTG GGCACCGGTG CATTTGTCAC	4500
TAGATACTTG CTCGGTAACC ATTCATCGCC GCAAGCAACG AGCCTTTATC GCCCTAGGAA	4560
GCAATATGGG AGATAAACAA GCAAACCTGA AGCAAGCCAT TGACAAACTG CGAGCTCGTG	4620
GCATCCATAT TCTCAAGAG TCCAGTGTCT TAGCGACGGA GCCTTGGGGT GGAGTGGAGC	4680
AGGATAGCTT TGCCAATCAA GTGGTTGAGG TGGAAACCTG GCTACCAGCA CAAGACTTGT	4740
TAGAAACCTT GTTAGCCATT GAGTCAGAGC TGGGACGGGT GAGAGAAGTG CATTGGGGAC	4800
CTCGTTTGAT TGATTTGGAC TTGCTCTTTG TGGAGGACCA GATCCTTTAT ACAGACGACC	4860
TCATATTGCC TCATCCTTAC ATAGCGGAAC GCCTTTTGT CTTGAGTCT TACAGGAAAT	4920
TGCGCCTCAT TTTATCCATC CGATATTAAA ACAACCGATC CGCAACTTGT ATGATGCTTT	4980
GAAAAAATAG AAAAATCTA GTTTTCAGTT ACTTGCAACT GAAGGCTAGA GTTTTATAC	5040
TCTTCGAAAA TCTCTTCAAA CCACGTCAGC GTCGCCTTAC CGTACTCAAG TACAGCTTGC	5100
GGCTAGCTTC CTAGTTTGCT CTTTGATTTT CATTGAGTAT TAAAATAGGT CATTTTCTTC	5160
TGGGAGGAGG ATAGTTTCTC TACCGTCCAT GTCTAAAACC AGTACTCTTG GGGGATAACG	5220
AGGGTCGAAA GGATGGTTAA ACTCAAAATC AATGGCTGTA GGGAGGTGTT GACTTGAAAA	5280
GTGGAAGGTA ATCTTTCCTT GOTTATTAAG CAATTGAAAC TCGAGTTCTT CTTCCAATTC	5340
AAAGACATTT TTAAGAAAT GGTCGATGAT ATACCAAAAA GAGTCAATGA TGTCATCAGG	5400

1298

CAAGCTGGTA ACAATACCAA AACTAGCAGA TCGCATGTGG GTATTGGTAA AAGCCATATC	5460
TCTGTCCCCT TTCTTTTCCC TTATCATACA GCAAATAGGA TAAAAATCA AGAAAAGGTG	5520
ATTTTTTGAA AAGGATTITA GTTACAGGGA GAAATAGGGA AAAAATTCCT AAAAATCTAC	5580
CGAAGTTAAT AGGTAAATTC CCAAATTAAC TTGATTATAT AACTTTCAGT TACTTTGAGA	5640
AGTTACCGAA AAATATTTT CATATCTATT GACTTTTAGG GGTAATAATTT GGTATGATAG	5700
TAGGCGGTAT TGTTTACCCC ATTTGAAAGG CCCCAGAAC TTCCAAATAC TTTTCGATGG	5760
GAAGGAACAC CCATCACCGT AAACAAAAAT CGAACTATAT ATAGGAGAAA TCATGAACAA	5820
AACAACATTT ATGGCTAAAC CAGGCCAAGT TGAACGTAAA TGGTACGTAG TTGACGCAAC	5880
TGATGTACCA CTGGACGTC TTTCTGCAGT AGTTGCTAGC GT	5922

(2) INFORMATION FOR SEQ ID NO: 268:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1988 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

TAATATCTA CGATGAGCTG TTGTGATTCT CATTAGTTCC CCTTTCCCAA GAGGCATAGG	60
GGTGCGCATA ATAGATGTGC TCCTCAGAAA ATATATCAAA CAAGCGATTG AATTCCGTTT	120
CATTATCTGC CGTGATGGAA AGAATCTTGT GTTGTMTTAA GATGAGTTTT AGAGCCTGAT	180
TGACCACCTC AGCACTTTTA TTTGGAATCA ATCGGATGAT CTGATGTCTA CTCTTTCGAT	240
CCGTCAAGAC AATCAAGCAG TAGTTTTTCG ATCTCGTAAG TAGAACCGTA TCAATCTCAT	300
AATGCCCAT TCTCAAGCGA AGATTGATAG CTTCAGGCCG CTGTTTCGATG GATTGACCAG	360
CAGGTTTAAA GTTGGTGCTA GCCTGTTTCT TAAGCGCTTT TCCTTTTCTA GGGTAAAGCA	420
AATCCTGCTT GCTTAACCCC AATTTTCCAT GATGAATCCA ATAGTAAATG GTTGAAATTC	480
CCACGTTAAC CCCTTTAGCC ATAACCATCA TTTCAGGCGA AAATTTTGG TTATGATAGT	540
GGAGAATCTT TTCCTTTAGT TCCTTGGTCA AGCTTGATTT CTTGACCGAG CGCTTGCGAT	600
TGTTTTCATA AGACTGTTGA GCGTAGTCGG CAGAATAAAC CTCTTTGAAG CGCCCTTTTC	660
CAAGACATTG TCGGACTGTC CCACGCTTGA TTTCAGTGTG ATAGTTTGAG GAGCTTTTCC	720
AAGTAGAGAG GCAATTTCTC TATTTGATTT TCCTTCTTTT TTCCATCTTT CGATTAAGCG	780
ACGGCTATCG ATTGTCAAAT GTTTGGCTTT TGTAGTATAA TTGTCTTGCA TCTCTGTGCC	840
TTTCTTGTGT TTGTGGTTGA ACAACAAGTA TAACACAGAG GTGCTTTCTT ATGCCTACAA	900

1299

GAGCTTTCAT TATTTCCATT TTCTTTTGGG TTTCACCTCTA TTCTGAAAAA CTTGTGTATA	960
TTTACTGAAG CTAGCAAGTC TTACCTGTAA ATTTAATGAA AGCAACACAA AATCCGAGAG	1020
GGGAATCTCG GATTAATAGA TAGAGAGTTT TTAGTTTAAA TAAATTGTTT AAAATATCAA	1080
CAACATCACT TCTTTTCTTA ACCTGATAAG TCTTGATTCC TAATTTTGGG GCTACGATTA	1140
TATTGTCTC AATATCGTCT AGAAAGACAC AATTCTAGG TTATAACTGG TATTTATCGA	1200
TAGTTACTCA TATACATCAG TCCACCTCCA TACTTATGTG CGAGCCTCTC TTTGTATTAT	1260
ACCTCCATAC TCACCTTACA GATTCTTTTG GTAATAATAT CTTTGCCTAA TGTAGAGACA	1320
GTCTTGCAAA GAAAAAACTT CTTGTAGCC ATGTTTCTGA TAAAAGTCCG GTGCCTGGAA	1380
CTGGTAAGTA TTGACAAAGG CAAAACAACA ATTCGATTG TTAGCTTCAC TTTCTGCCTG	1440
TTGCAATAGT TTTGAACCGA TTCCTTGCCC TCGCAGTTCC TCTTTTACAA ACAAATACTC	1500
GATTTCTAGC CAATTTCCAA AAGTCTCTGC TATCAAACCT GCCAGGAGAT TGCCCTTTTC	1560
ATCTTCGACA TAAAGATTAA GTGGCTCACT TTCAGCCTCT TCTCTTTTGG AACGGTTATA	1620
AACACGAATC AGATTCCCTA TTTCTTGCGA TTTATGTGAT TCCTTATTTT CCAATCTAAA	1680
GTATAGTGAA ATGAAATAAA ACATGCGCAA ATCGATTAAG GAATTTAATC TAATTTCTAA	1740
CAATGTCTTA GAAATCAAAG TGTACTATTT TAACTTCAAT GCACTATACA TCTAATACTC	1800
AATAAAAAATC AAAGAGCAAA CTAGGAAACT AGCCGCAGGT TGCTCAAAAC ACTGTTTGTG	1860
GGTTGTAGAT AGAACTGACG AAGTCAGCTC AAAACATAGT TTTGAGGTTG TAGATGAAAC	1920
TGACGAAGTC GGCTCAAAAC ATGGTTTGTG GGTGTAGAT GAAACTGACG AAGTCAGCTC	1980
AAAACAGG	1988

(2) INFORMATION FOR SEQ ID NO: 269:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 709 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

CCGGATATTT GTTTTATGTA ATTTTCTTGC AAGTTTCTTC TTAGTAGCTT GTCAGTCAGG	60
TTCTAATGGT TCTCAGTCTG CTGTGGATGC TATCAAACAA AAAGGGAAAT TAGTTGTGGC	120
AACCAGTCCT GACTATGCAC CCTTTGAATT TCAATCATTG GTTGATGGAA AGAACCAGGT	180
AGTCGGTGCA GACATCGACA TGGCTCAGGC TATCGCTGAT GAACTTGGGG TTAAGTTGGA	240

1300

AATCTCAAGC ATGAGTTTGG ACAATGTTTT GACCACTCTT CAAACTGGTA AGGCTGACCT	300
AGCAGTTGCA GGAATTAGTG CTAAGTACGA GAGAAAAGAA GTCTTTGATT TTTCAATCCC	360
ATACTATGAA AACAAGATTA GTTCTTGGT TCGTAAGGCT GATGTGGAAG AATACAAGGA	420
TTTAACTAGC CTAGAAAGTG CTAATATTGC AGCCCCAAAA GGGACTGTTT CAGAATCAAT	480
GGTCAAGGAA CAATTGCCAA AAGTTCAATT AACTTCCCTA ACTAATATGG GTGAAGCAGT	540
CAATGAATTG CAGGCTGGAA AAATAGATGC TGTTCATATG GATGAGCCTG TTGCACTTAG	600
TTATGCTGCT AAAAACGCTG GCTTAGCTGT CGCAACTGTC AGCTTGAAGA TGAAGGACGG	660
CGACGCCAAT GCCGyTGCTC TTAGAAATA GTGATGATTT GAAAGAAGT	709

(2) INFORMATION FOR SEQ ID NO: 270:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1680 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

TATAAAATGT TAAGTTAAAT GATTTCAAAA TTCAGAAAGG GATTGCTTTA TGCAGTTCCT	60
TTTTATTTTA ACAGGAGTGA AACTATAGTG TTTCTAAATT GTGAATCAAT CAAAAGTATG	120
TGTGATGGGG CTATTCTAGC TTTAGAAACC TTCAAAAATT AAAATTTAAG GCAATCAATT	180
ACTTGAAGA GTATGAAAGC ATTTAGTTTA TAGGAATTCT AGGTCTAGAA TTACATATAT	240
ATATTTATGA AGACGGGGTG TTCGATAGTT AGTATTGTTT TATTCTGAAA GATTTGAGCT	300
GTCAGTTGTA TAGAAAGTGT TCGAATTTTT TTAAGTGATT AAATTAGTTA ATTGTATGAG	360
GTGCTTTATG ATATAATGTT CTTAATGAAT TTTCAGAAAG GAAAACCTCA AATTGTTCTA	420
CAAATTTCTA CTCTTCGACC TCGACCACAC TCTTCTTGAT TTTGATGCTG CTGAGGATGT	480
GGCTTTGACC CAACTTCTAA AAGAAGAAGG AGTTGCGGAT ATTCAGGCTT ATAAAGATTA	540
TTACGTTTCT ATGAACAAGG CTCTCTGGAA AGACTTGGAG CTGAAGAAAA TCAGTAAACA	600
AGAGCTGGTT AACACGCGCT TTTCTCGTTT ATTTGCTCAT TTTGGACAGG AAAAAGACGG	660
TAGTTTTCTT GCCCAGCGTT ACCAATTTTA CCTCGCCCAG CAGGGACAAA CACTATCGGG	720
CGCTCATGAT CTCTTGGACA GCCTCATTGA GCGTGATTAT AACTTGTATG CTGCGACAAA	780
TGGCATTACT GCCATTCAGA CAGGACGTTT GGCTCAATCT GGTCTAGCAC CTTATTTCAA	840
TCAAGTCTTT ATCTCAGAAC AGTTGCAAAC TCAAAAGCCG GATGCTCTTT TTTATGAAAA	900
GATTGGCCAG CAAATTGCTG GATTTAGTAA AGAAAAGACG CTGATGATTG GAGATTCTCT	960

1301

AACCGCCGAC ATTCAAGGTG GCAATAATGC GGGGATTGAC ACTATCTGGT ATAATCCTCA	1020
TCACCTCGAA AATCACACAC AAGCCCAGCC GACTTACGAA GTCTATTCTT ACCAAGACTT	1080
GCTGGATTGT TTAGATAAAA ATATTCTTGA AAAGATCACA TTTTAAAGGA GACGAGCTAA	1140
TGACTACAAA AAAGCTAATA TTACTATTGA AGAGTACATT GAAATGTCTG AAGTTGATTT	1200
TAATGAAGCT GTTAATTATG AATTACATC TGACACTTGT CAATTAGCAA ATAGTATTTA	1260
TCAATCTCTT TTTAAGTTT TTGATAAGAA AAATTTCTCT GCGGATTTAA TTTTACTTG	1320
GAAATCTCCA TCATTAGTCA AAGAAGGGA TTATATTGGG AGAAGGGATT CACAAGTAGA	1380
TAATCTTAGA GTAATAGGAA ATATATTTCC GAATTATCTT ACTAATCGAA AATATAGCCT	1440
CAATATGAAT CGTAATGGCT GTATGGGAGA TTTTCCTCAT GACTTTTTTG ATATATACCT	1500
AGATCATGTA GCAAAATATG CCTACGAACA AAAAGTTAAT AATATTAAAG AGTATTATCC	1560
TTTAAAAGA GCGATTTTAC ACCAAGAGAA TGCATTGTAT TTTCGATTTT TTTCTAATTT	1620
TGACGACTTT TTAGAAAAA ATTATTTAAA GACTATATGG CAAGTTTCTA AAGAACTCC	1680

(2) INFORMATION FOR SEQ ID NO: 271:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 598 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

AGCTCGGTAC GTAGTATnTG TGGTGCATAA ATGAGTGAAA AGAGGATAGA GAGGATGAGG	60
CCGATAAGAA CACCGGTAGC TGCATCGTGA AATACTTGTT TTTTCATAGT TCTAATTTCT	120
CCTTGATGGT TTTTAGATAA CGGCGTGAAG AGTAGGTGAA GCTTTCGTTT TTCAAGAAAA	180
TTTCTACCAG ACCGTTTGGC GTGA <sub>g</sub> CTTGA GGTGAGAGAT GGAATCGATA TTGATGATTT	240
CTGATTGGGA AATTGCGATA AAATTGGTTG GCAAGAGTTT AAGAACCTGA TAGAGTCGCA	300
AATCAATGCT GTAGGTCTGA CTCGCGGTTT CTGCTAGAAC CTTCCGATTC TCGATATAGA	360
AGCGCTGAAT CTTGCCAATC TCAACTAGAT AGACCTGATC ATCGATTTTT CTTTGATTTT	420
TTTCTCTTTG GTCCAGATTT TCTGCGAACT CGATGACTTT CTGGACTTTT TCGGTTTCTT	480
GAGGTGCTTG GACAATCAGC TTTTCCTCCT CGTAAGTCTC ACTAATCTGT AGTTCTACTT	540
TCATAGTTTT CTCTCCTTTT CAGTTATACA AGGTTGTGAT CACTTCCTGT ATATCCGG	598

(2) INFORMATION FOR SEQ ID NO: 272:

1302

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1099 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

```

CCAGCAAATC AATAACTGCA ATTGCTATAA AATGGATTCT ATAGAGTTTT TTCATGACAA      60
GACCTCCCTC TTTTATCTAA CTTCATTCTA CTCCAAAAGA ATGGGAGTTA CAACTAAAAT      120
GATAAAAATA GCAGAAGGGA GATTCTCTTA AGTTGGCTAG TATTCTTTAT TTGAGTTTCC      180
TTCTATTATC TAACTTCTTC ATCATTCCAG ACAAATAAAG CTCCGATTGC ATTGAGGATA      240
TAAAAGATGT ATTTACCGAT ATTGGCGAAG TTTCCTTGAA TACCAGCTTT TGTCAGCTGA      300
ACGAAATGTG AAATCAACCA AAAGCCCCAC TGAGTTGTTA GTTTTAATGC ATTCAAAGCA      360
TTGGCAATGA GGGACAGTGC AAAGGCAATA GTTGTTACGT AGGCAAGGAG ATTCATCTTG      420
CCCCCATATC CGATATAGTT GGTCACAAAG GCAAAGAGGA AGGCGATGAT GGAAATGATG      480
ATGGCCGCCA ATTTTACCTG TTTTGGCTC ATTTGGTTGG GTCTGCCTTC TTGCGAAGCT      540
TCCCACTTCT TTATAGCAAA GGTATAAATG AGGAAGGTGA CGGGATAGGT AATGATGGCC      600
GCCTTATTTC CAAGGATATA ATCAATAGCA CCGGACAAAA TGGTATTAAC AATACCAAAG      660
TAATTTCCCC ATTTGCTTAA TTTCCCCGTG AAACGAGTGG ACAACATGGA AATCCCAACG      720
TTGGTTACGG AAATCAATCC AAAGGGTACA AGAGCTGTCC ATGATCCCCA GTCTACAAAT      780
TTATCGAGGT GTGAGTTGAG GTAACCAGAT GCAATCGCAA TCCCAACGAC CAAAGCAACC      840
CCGAAGAGGT CAACTATTTT AGATGTAGCA AAAATTTTTA GTGATTTTTT CATAGGTTAA      900
ACTACCTTTC TTTTTTCAA ATATTCTCCC ACCAAATGAA AGTAAAATAA AATGATAGAA      960
ATAAAACCCT GAAAATAAAG GTTCTATAAT ATTTGTAGTG GGTAAATCCA CTATAGATAT     1020
TATGGAGCCT ATTTTATTGT AGAAAAAAG TCCCATATGA CCTATAATGA AAAGCGACAA     1080
AACAACTCAT TAGAAAGAT                                     1099

```

(2) INFORMATION FOR SEQ ID NO: 273:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2723 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

1303

CTGGGATTCA CGTGAAAAGG AAGCCCAGAG AGTAGCCAGG TGTACTGCTA GAACAGTGAG	60
TGAAATTGAA TATTACCATA GAGAGTCAAC CCAGATAGCT CAGGCTTTAG TTGAAAATCA	120
AGCTCGTATC GAGGGAATCT ATAAATACTT TAGCCTTAGC ATGCCAGACT ATTTTTACTG	180
GCAATTAGAG CGGAAAGCTT CGCCTTATAT ATCAGTCTCT CTGTATGAAA ATGTTGATGA	240
CCTCTATGTT CGAAATGATT TTGTAAGTGG GGTGGCCATT GCTTTTCAAG ATTACAAGGA	300
AGTCTATGTT TCTACTAAAG ACAACGCTAG GKKAGAAAAA ATCAGGGCTG AGGATTTCAA	360
ACCAGCAGGA AATAGTTTTG CCATTCCAGT GTCAGATCCA GTGTCAGATC AAGACTTAGG	420
AGTGATTTAC ATCTCCTTGG ATCCTGCTGT TTTATACCAT GCCATTGATA ATACTAGAGG	480
TCATACTCCG ATGGCAGTAA CAGTGACCTC ACCTTTTGAT ACGGAGATTT TTCATATGGG	540
TGAGACAGTT GATAAGGAGA GTGAAAATTG GCTAGTTGGC TTAAGTTCTC ATGGATATCA	600
GGTTCAGGTG GCAGTTCCTA AAACTTTGT TTTACAAGGA ACAGTGACTA GCTCTGCTTT	660
GATTGTGGGT TTGAGCCTTC TCTTTATGT CATCTTTAT CTGACTTTGA GGCAGACTTT	720
TGCTAATTAC CAAAAGCAGG TAGTGGATTT AGTAGAATCC ATTCAAGTCA TTGCTCAAGG	780
CGAAGAGGGG CGTCGGATTG ACATTCCGA GAAAGATCAG GAATTACTCC TAATCGCGGA	840
GACGACCAAT GATATGTTGG ATCGATTGGA AAAGAATATC CATGATATTT ACCAGTTAGA	900
GCTTAGTCAA AAAGATGCCA ATATGCGAGC CTTCGAGCG CAAATCAATC CTCATTTTAT	960
GTATAATACG CTGGAGTTCT TGCAGTATGA TGCAATTATG CAGAGTCAAG ATGAGTTGGC	1020
AGATATCATT TATGAATTCA GTAGTCTCTT GCGTAACAAT ATTTCCGACG AAAGAGAGAC	1080
CCTCCTCAAA CAGGAATTAG AATTTGCGG TAAATACAGC TATCTCTGCA TGGTTCCGTA	1140
TCCCAAGTCC ATTGCCTATG GTTCAAGAT AGATCCAGAG TTAGAGAATA TGAAGATTCC	1200
CAAGTTTACC TTGCAACCGC TGGTAGAAAA CTATTTCCGG CATGGTGTG ACCACAGGCG	1260
GACAGATAAT GTGATTAGCA TCAAGGCTCT TAAACAGGAT GGTTTTGTGG AAATTTTGGT	1320
GGTCGATAAT GGTAGAGGAA TGTCGGCTGA AAAGTTGGCA AATATCCGAG AAAAATTAAG	1380
TCAGAGATAT TTTGAACACC AAGCCAGCTA CAGTGATCAA AGGCAGTCTA TCGGGATTGT	1440
CAATGTACAC GAGCGTTTTG TGCTCTATTT TGGAGACCGC TATGCCATTA CTATAGAGTC	1500
TGCAGAGCAA GCCGGTGTTC AGTATCGTAT TACAATTCAA GATGAGTAGA AAGGGAGAAA	1560
ATGTATAAAG TATTATTAGT AGATGATGAG TACATGGTGA CAGAAGGTCT GAAGCGTTTG	1620
ATTCCTTTG ATAAGTGGGA TATGGAGGTC GTCGCAACAG CCAGTCATGC CGATGAAGCT	1680
CTAGAATATG TTCAGGAAAA TCCTGTCGAT GTCATCATTT CCGATGTCAA TATGCCAGAC	1740

1304

AAAACAGGGC TTGATATGAT TCGGGAGATG AAAGAGATCT TACCAGATGC TGCCTATATC	1800
CTGCTCTCAG GTTATCAGGA GTTTGATTAT GTAAAAAGAG CAATGAACCT TAGTGTGGTG	1860
GACTATTTGG TCAAGCCTGT TGATAAGGTA GAGCTGGGAA ATCTGCTGGA GAAGATTGCA	1920
GGTCAGCTCG GCGAGAGAGG GAAGAAAAGT CAGACTCTTA GTCAAGAATT AGACGAGGCT	1980
GGATTTGTTA GTTATTTAGG GGATAAGGAG AATTGGTGGA TAGGTCTATC CAAGGAAAAA	2040
CAAGGTTCCCT TCACCATTCC CTACTATGTC TTGGGTCAAG ACTGGCAGAT TTTCATTCT	2100
GGCCACCCCC TAGATGGTTT AGTCGTTACA CCTTTTGAAG CTCCTTATCA AGAACACTTT	2160
GAACGCTGGA AGCTGAATGC TGAGAAAACC CTCTTTTACG GTTCTGTAAA TCTGCAGCAG	2220
TCTGAGAGTC TCTTTGCCA TTACGAACCG ATTTATAGGG TTATCATPCA GGGAAATCTC	2280
AATCAAATCG TAGAAGAGTT AAATCTCTTG GAGAAGTAG TTCTGAAAA TACACCTCGT	2340
GTTCGATTCTA CTAACAGCT TTTTATCCAG TTTGTCATGG ATGTTTCCA TTTATTTGAA	2400
CATCTCAAAG CTGATGATAT GACGGACATT GTCAAAACCA TTCATGCTAT TCAATCCTTC	2460
GATGAATTGG TTTCTTATAT CAAGGAAACT CTGATCAGCT TTTTCGGTCA ATACCGTATG	2520
AATGAAAATG TGGTCAGTGT GCTGGAAGTC ATTGGTCGTG ATTACCAAAA AGAGCTTTCC	2580
CTCAAGGATA TCAGTAAGGC CCTCTTTATC AATCCTGTCT ATCTAGGGCA GTTGATTAAAG	2640
CGTGAAACCG ATTCGACCTT TGCAGAGTTA CTAACAAAC AACGTATTAA GGCTGCCCCAG	2700
CAGCTCTTGC TTTCAACTAG TGA	2723

## (2) INFORMATION FOR SEQ ID NO: 274:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 836 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

CCGCAGTTTT TTAAACCGT ATATAAGTAT AGCATAGTCA AAAAAAGAAT GCAAGATTTT	60
TGCAAACTTT TTTAAATTT TTCGTAATTT TTCTTTTAAA GTTCTACTGT CAGGACTTGA	120
CCTTGCTTAA CAACCTGTTT TCCGGCGATA TAAACATCAT CTACATCACT AGATTTAACT	180
GCATAAACCA GGTGAGACAG CATATTTTCC TGAGGTTGGA GATGAATTTT CCCTTGTTGGT	240
TGAATGACCA GAAATCTGC TTGCTTGCCG ACTTCCAGAC TTCCTATCTG ATTTTCCATT	300
CCAAGGACCT TAGCCCTTC GATTGTCAGT ACCTTGAGAG CTGTTTCGAT TGGAAACTGG	360
CTGGCATCCC CACTTTTCAT CTTCTGAAGA AGAGCTGCAG TCCTTCCTTC CTCAAACATA	420



1305

TCTAGATTGT TATTGGAAGC AACCGAGTCA GTCGCAATTC CGACTGCTAC TCCCGCTTTT	480
TGGAGCTGGA TAATTGGAGC AATTCCTGAT GCCAGTTTGA GGTTACTGAT AGGATTGTGG	540
GCGATAGCnA CTTGAGAAGA TGCCAAGCGT TCAATTTCTC TCTCGTTTAA TTCGACCCCG	600
TGAGCAAATA CGGACGGATG ATCTAAATAA CCCAGTTCTT CAAGAAAAGC AAGGGGGCGT	660
TTGCCGTATC GTTTGAGGAT AATTCCTGAC TCCTCCTTGG TCTCCGCCAC ATGGACATGG	720
AGCGGAATAT TTAGCTCTTT TGCCATTTC AACTCGCTT CCAGCAAGTC TCTACTGCAG	780
CTATACGGAG AATGAGGTGC TACCATAACC TTGAAATTG GATTTTATA TTTTAA	836

(2) INFORMATION FOR SEQ ID NO: 275:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2335 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

ATTTTATTC ACTTTTAGG TGGTCTGGG CTATTCTTAT ATAGCnTCAA GACCATGGGA	60
GACGGTTTAC AACAAAGCTGC TGGAGATCGC CTGGGTTTTT ACATTGACAA ATATACTAGT	120
AATCCTTTGT TTGGAGTTCT GGTTGGTATT GGGATGACTG CTCTAATCA GTCTAGTTCT	180
GGTGTAACAG TTATCACAGT CGGCCTGGTC AGTGCCGGTC TCTTAACCTT ACGTCAGGCT	240
ATCGGGATTG TCATGGGTGC TAATATTGGG ACAACTGTCA CATCCTTTCT CATCGGTTTT	300
AAATTAGGTA ACTATGCCCT ACCTATGCTC TTTATCGGTG CCGTCTGTCT TTTTTTACG	360
AAAAATCGGA CAGTCAATAA TATCGGACGC ATCCTCTTTG GTGTCGGTGG TATCTTTTTT	420
GCCCTCAATC TCATGAGCGG CGCAATGGCT CCACTCAAGG ATTTACAGGT CTTTAAGGAC	480
TATATGATTG AGCTAAGTAA GAATCCTGTT TTGGGTGTCT TTGTCGGTAC TGGCTTGACC	540
TTGCTAATTC AAGCTTCTTC GGCTACCATT GGGATTTTAC AAAACCTCTA CGCCGGCAAT	600
CTAATTGATC TACAGGGAGC TTTGCCAGTT CTATTGGTG ACAATATCGG GACAACCATT	660
ACAGCCATCA TTGCCTCTTT AGGGGCTAAT ATTGCAGCTA AACGGGTAGC AGGAGCTCAT	720
GTTCCTTCA ACGTTATCGG AACAGTTGTC TGCCTTATTT TTCTAGTTCC TTTTACTGTC	780
CTGATTCATT GGTTTGAAGC TACGCTAAAT CTAGCACCGG AAATGACCAT CGCCTTTGCT	840
CACGGAACCT TTAATATTAC CAACACCATT GTCCAATTC CATTTATCGG AGCTCTGGCT	900
TACTTTGTAA CCAAGATTAT TCCTGGAGAG GACGAGGTG TCAAATACGA ACCCTTATAT	960

1306

CTTGATGAAC ATTTTCATCAA ACAGGCCCCA TCTATCGCTC TAGGAAATGC TAAGAAAGAG	1020
CTCTTGCACT TAGGAACTA CGCTGCTAAA GCCTTTGACC TTTCCTATAA GTACATCATT	1080
GACTTGATG AAAAAGTTGC TGAAAAAGGG CATAAAACCG AAGAAGCAAT TAACACCATC	1140
GATGAGCAAT TAACACGTTA TCTCATTGCC CTTTCAAGCG AAGCTCTCAG CCAAAAAGAA	1200
AGTGAAGTGC TTACCAATAT CCTTGATTCC TCCCGTGATT TGGAACGGAT TGGAGACCAC	1260
ACGGAGGCTC TACTCAATCT GACTGACTAT CTTCAACGGA AAAATGTTGA ATTTTCTGAT	1320
GCCGCCTTGA AAGAATTAGA GGAAGTTTAC CGCCAACTA GTGACTTTAT CAAAGATGCT	1380
CTGGATAGTG TGGAAAACAA TGATATTGAA AAAGCACGCA GTCTTGTA AGTCATGAA	1440
GCAATCAATA AGATAGAACG TGTTCTCAGA AAAACCCACA TCAAACGCCT CAACAAAGGC	1500
GAATGTTCAA CACAAGCTGG GGTCAACTTT ATCGACATCA TCTCACACTA CACTCGTGA	1560
TCAGACCACG CTATGAACCT TGCTGAAAAG GTTTTTGCAG AACAAATCTA AGAACCAAGA	1620
AGCTATCCAT CATAATTGGA TGGCTTTTTA CTTTTTCTA AGCAAGACTA GGATGAATGA	1680
AACTGAAAGA GTATTCTGCA GATATATAGT CCCCAATTAT TCACCCCAA TCTAAAAACC	1740
ATCCAGAATC CTTGCCCTAG CTTAGATCCT GGATGGTTT TTTTTCACC CAATGGGTGT	1800
TTTTTACTAG ACAAAAAAGA GTTTCCCTT TATGGTATAA GTGTAGAAAA AAACACAAAA	1860
AGAAAGGAAA CTCACATGAA CAGTTTACCA AATCATCACT TCCAAAACAA GTCTTTTAC	1920
CAACTATCTT TCGATGGAGG TCATTTAACC CAGTATGGTG GTCTTATCTT TTTTCAGGAA	1980
CTTTTTTCCC AGTTGAACT AAAAGAGCGG ATTTCTAAGT ATTTAGTAAC GAATGACCAA	2040
CGCCGCTACT GTCGTTATTC GGATTGAGT ATCCTTGTC AGTTCCTCTT TCAACTGTTA	2100
ACAGGTTATG GAACGGACTA TGCTTGTA GAATTGTCAG CTGATGCCTA CTTTCCAAAA	2160
TTGTTGGAAG GAGGGCAGCT TGCTTCACAG CCAACCTTAT CCCGTTTTCT TTCCAGAAT	2220
GACGAGGAAA CAGTCCATAG TTTGCGATGC CTCAACCTTG AATgGkCGAA TTCTTTTAC	2280
AGTTTCACCA GCTAAACCAA CTCATTGTAG ATATCGATTC TACCCATTTC ACAAC	2335

(2) INFORMATION FOR SEQ ID NO: 276:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 752 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

CGGATTCACCT GTTGTGACT AATCAATAAC ACAGTAGAAA ATCTCACAGC AGTCTATTAG	60
---	----

1307

TTGCTTTTCA TACTAGGCAA GTGACTGAGG CTTGTACTTG GGTACAGCAA GGGAGCTTAA	120
GGCCGTAGAA GAGAAAAATA GTAGACTGAA AACCCGCAAG ACTTCATCAT TTCGAGAAGT	180
GACGTGGGAG ATGAAAATCG ATTGAACCAC TTACAAGGAG AATAGAAAAT GGCTAAAAAA	240
AGCAAACAAC TTCGTGCTGC TCTTGAGAAA ATCGACAGCA CAAAAGCATA CAGTGTAGAA	300
GAAGCTGTAG CACTTGCAAA AGAAACTAAC TTTGCAAAAT TTGATGCAAC TGTAGAAGTT	360
GCTTACAAC TGAACATCGA CGTTAAAAAA GCTGACCAAC AAATCCGTGG AGCAATGGTA	420
TTGCCAAACG GTACTGGTAA AACTTCACGT GTTCTTGTTC TCGCACGTGG TGCAAAAGCT	480
GAAGAAGCAA AAGCTGCTGG TGCAGACTTT GTTGGTGAAG ATGACCTTGT TGCTAAAATC	540
AACGACGGTT GGTGGACTT CGACGTAGtT ATCGCTACAC CTGATATGAT GGCTCTTGT	600
GGACGTCTTG GACGTGCCT TGGACCACGT AACTTGATGC CAAACCCTAA AACTGGTACT	660
GTAACAATGG ATGTTGGCAA AGCGGTTGAA GAGTCTAAAG GTGGTAAAAT CACTTACCGT	720
GCTGACCGTG CAGGTAACGT TCAAGCAATC AT	752

(2) INFORMATION FOR SEQ ID NO: 277:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2643 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

GTCAACATTG ATTTCAAGGC TGTGTGCTTT CTATCTCCCC TTTTTCATAA TGTATAATAA	60
AATGAAATAA TAACAGGACG AATTGATCGG GACAGTCAAA TCGATTCTA ACAATGTTTT	120
AGAAGTAGAG GTGTACTATT CTAGTTTCAA TCTACTATAT TTTGCTACAG GTGCTTCAAC	180
CATTTGAACG ATTTCAAATC CTTCTTTTGG GTAAAGATTC TGAGCTCTTT GATTGCTCTC	240
GAAGACATTT AGAGAAATAC TGTCTATATC TCTATTTTCA AATGCTAAAC TAACAAATTT	300
CCTTAAAGCC TTGCTACCTA AGCCTTGCTC CTGTTTCTGG GGGTTGATAA AAAATCTCCC	360
GATATGAAGA TTGCTGTCTT CTAGCCTGAT TTTCTGGATA AATCCCACAA ACTCTTGTTT	420
ATCAAAGATT GAAAAGACTC CTTCCAAGGC TTGAAGTGTC AGTAGAAAAG GAATCCTTGG	480
TCCCATCCAT TGTCTTGAA AGGATTTGCC TAGGGAGTTG GACCACTGGC ATACAAATTG	540
AGCGTTTTCT GTGCTCACCT TTTCTTCAA ACGAATTGTC ATCTTTTCTT CACCACCTTA	600
TCTATGTTTC TCCATTATAC TATTTCTCCC ATTTTTTACG AATAGATAAG TATGATTGAT	660

1308

TTTATTATTT TTCTCGTCGG GAGCATTCTA GCTTCCTTTC TTGGTTTGGT CATTGACCGT	720
TTTCCAGAGC AATCCATTAT CAGTTCAGCC AGTCACTGCG ATTCCTGTCA GACTCCCTTG	780
CGTCCCTTAG ATTTGATTCC GATTCTCTCA CAGGTCTTCA ATCGCTTTCC CTGTCGCTAC	840
TGCAAAGTTC GCTATCCTGT CTGGTATGCC CTCTTTGAAT TAAGCTTAGG ACTCCTCTTT	900
CTGCTTTACT CTTGGGGATG GCTCTCCTTG GGGCAAGTCG TCCTAATCAC CGCTGGTTTG	960
ACCTTGGGTA TCTACGACTT TCACCATCAG GAATATCCCT TACTGGTCTG GATGACTTTC	1020
CAGCTAATCC TAATAGCTTC CTCTGGCTGG AATCTGGTCA TGGTCTCCTT CCTCATACTT	1080
GGAATTTTGG CTCATTTTAT CGATATCCGC ATGGGTGCAG GGGATTTTCTT CTTTTAGCT	1140
TCTTGTGCTC TCGTCTTTAG CGTAACGGAG TTAAGTATCT TGATTCACTT CGCTTCTGCG	1200
ACGGGTATCC TGGCCTTTCT CCTGCAAAAG AAAAAGGAAA GACTTCCTTT CGTGCCTTTC	1260
CTCTTACTTG CTACTTGTTC GATTATTTT GGTAAGCTAC TGCTTGTCTG ATAAATCCA	1320
ATTTCTGCCA TATATCCTTC ATGAAATTAT TTCACAGTTA AATTATAAAT TATTTCTTTT	1380
GTACAAAGGG ATGATGTTAT CAAATCGATC TGTTCTTCTA TCTTCTTGAT ACTGATCAAA	1440
AAATTTCAAT TCGACTGAAA ATATTTCTGCT TATAAACTGT AAACGAATAC TTTGTTTAGA	1500
CATTATAGTC GCTAGACTGA CTAGATGATT ACTCAAAACG ACGTCCAGAA TACTCTTTAC	1560
TTTGCTTGGT TTTTAAACAA AAATTTGATC ATCCAAGGGT TCAATCATTT TGTAACCTTT	1620
TTGCGCAATT TGACGATAAA AGTAAGAATG TTGCTTTGGA GTCAATAATC CTAACCTAAA	1680
AGCTCGATAC TCTAAAGCCT GTATCGAAAC ATTCAAATCC GACTTCAATA AAATATAACT	1740
ATCAGGATTG CTGACACGCT TGCCAACCTT CTCTTCAAAT TTGACTAAAA ACTCTTCTTT	1800
TGGCAATAAA AAACATGATG CAAAATAATT TGCTTCTTGC TCCAAACGAT CGCCATCTTC	1860
ATTCATATCT TTATATTTAT GTAAAAGAAT ATGTCCTAGC TCATGAGCTA AGTCAAAATT	1920
TCGACGTACA GATGATTTAT TCGTTCCTAA CACAATATAA GGTCTTCCCA ATTTTGACCA	1980
TGCGCTATAA GCATCAGCTT GGCCATTAAT TAATCGTTCC ACGATATAGA TGCCTGAACG	2040
TTCTAATTTA TAAAGCAAAT CATGATTATC TTTTGAAATA CCTAATTTTT CCCTGGCATA	2100
AAGAGCCAAT TCCTCAATGG ATTCTCCCTT ATGATAAGAT TCACTCACTA CATTACTTAG	2160
GTCAATGAAT ATAATATTAG GTATAATTAC AAACTTTTCA AAATAATCAA TCAAACTATC	2220
TACCTTATGT AAATACATAG TTTGAATATC TATGTTTTTC CGTGTGCTA GGTCTGCATT	2280
TCTAAAGGCA ATTACAGAAG AATCAAATCG AATGCTCTCT TCTTCTGTT CAAAATAAGT	2340
TAAATCAACA TGAAATTGGT TGGCCAAATG CATTTTGGTT GATAATTTAG GTTTCGTTTC	2400
GTTGGACTCA AACTGCCAAA TGGCTTGTTT CGTTAAATTA ATTCTCTGAG CTAATTCTGC	2460

1309

TCTACTTAAA CCATTTAACA GCCGTAATTC TTTCAATACC CGACCATTAA ACATTTACAT 2520  
 ACTCCTTACT ACTTTTGACC TTCTTGTTTT TCTATTCTTG GAATAATTC AAAATCTTCT 2580  
 GTTTCGGATA ATTCTGAAAA ATTAGGAATA TCTTGATATT TAGCTTCTTC GAAATGGTAC 2640  
 GGG 2643

(2) INFORMATION FOR SEQ ID NO: 278:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 582 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

TGACCACTGG CAAAATGGCT ATCCAAATGC AGATGTTATT ATCGATGATA TCATCTCAGG 60  
 GCAAGCCTAC GTAGCCTTGG AAGAGGGAGA ACTGCTAGCC TATGCTGCTG TGACCAAGAG 120  
 TCCAGAGGAG GCCTATGAAG CTATTTATGA GGGAAACTGG CAAGCTGGAG AGTCAGAGTA 180  
 TCTAGTCTTT CACCGTATTG CTGTGGCAGC AGATGTGCAG GGAAaAGGAG TTGCTCAAAC 240  
 CTTCTTAGAG GGCTTGATTG AAGGTTTTGA TTATCTTGAT TTTCGCTCAG ATACGCATGC 300  
 TGAAAACAAG GTTATGCAAC ATATTTTGA AAAACTTGGT TTAAACAAG TCGGTAAGAT 360  
 GCCAGTAGAT GGCGAACGCT TGGCCTATCA AAAATTAAAG AAATAATGCA AAAGAAGTAT 420  
 GTAAAAATCC TCTACTCCTC ACCAATTGGT ATTCTATCAC TTGTAGCTGA TGACCATTAT 480  
 TTGTATGGAA TTTGGGTTCA GGAGCAGAAG CATTTTGAGA GGGGACTAGG AGATGAAACG 540  
 ATAGAAGAAG TTGTWAGTCA TCCTATTTTA GACCCAGTTA TT 582

(2) INFORMATION FOR SEQ ID NO: 279:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 554 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

CCCAAGCTAC TAAGAGACTA AAAGTTGCTA GAGAAGCAAG AGAAAGTGTG AATCTTTTAA 60  
 ATTTATGAT GAATTTCTTT TCTGCTACCA ATTTAGAGAA ATTTTCTCTA ACCAGCAATT 120  
 CCCCTAGTAT AACAAGTTCA AAAAATGGAG TCAATTTATC TGCTCACGGT CCAGCAGGTA 180

1310

GCCCCGTACT TCTGAGATAA AATAGAGAGA CCCTGTAACG AACAGCAAGT CTTGAGCGTC	240
TGCCCTTTCT TCAAAATCGC TGATAAATTC TCGGTAAGAA GAACTATAT CGTAACCTGT	300
CACATCCCTT TCGTCCAAAG CCCCCTGATA GTCAAAGCCG GTCACCTTGA GTTCCACCTG	360
AGGCAATTTT TCAGTCAGAT AACCCAACAT CCCTTGATAA TCCTTACGTT TCAAGGATCC	420
AAAGAGGATT TGAGGTCGAT AGCCTTCCTG CTCTTTTCTT TTGATAAACT CAGCCAAGCG	480
AGTCAAGGCA GGGAGGTTAT GAGCACCATC CAAATAAATC TGTGGGCGAA TACGCTCCAA	540
GCGAsCAGCC CAAT	554

(2) INFORMATION FOR SEQ ID NO: 280:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 766 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

CCGGTTTTTC AAATGAATTT CTTGGTTGTG GCTAAAAAAT ATGCTACACT ATCAATATGA	60
AAATTTTAAT CCCAACAGCA AAAGAAATGA ACACAGACTT CCCAAGTATC GAGGCAATTC	120
CTTTAAACC AGAAAGTCAG GCCGTGCTTG ATGCCTTGGC TCTCTATTCT GCCAGTCAAT	180
TGGAGAGTTT CTACAAGGTA TCAGCTGAGA AAGCGCGGA AGAATTTCAA AATATCCAAG	240
CTTTGAAAAG GCAAACTGCT CAACACTATC CAGCCTTGAA ACTTTTTGAT GGGCTTATGT	300
ACCGCAACAT TAAGAGAGAT AAGCTGACCG AGGCGGAACA AGATTATCTT GAAAATCATG	360
TTTTCATTAC CTCGGCTTG TACGGTGTG TTCCAGTCTT GTCACCCATG GCTCCTCACC	420
GTTTGGATTT TTTGATGAAA TAAAAGTCG CTGGTAAGAC TTTGAAGAGC CATTGGAAGG	480
CAGCCTATGA TGAACTCTG AAGAAGGAAG AAGTGATTTT CTCTCTCTTG TCATCAGAGT	540
TTGAGACTGT ATTTTCTAAG GAAATCAGAG CAAAGATGGT GACCTTCAAA TTCATGGAGG	600
ATAGAGGCGG TCAGCTGAAG ATTCACTCAA CTATCTCCAA GAAAGCGCGC GGGGCCTTTC	660
TAACAGCTTT AATAGAAAAT CAAGTACAAA CTGTGGGGG AGCACGTCGC TTGAACTTTG	720
CTGGATTTGT TTACCGAGAA GATTTGTCAC AACCACAGGG GGATGG	766

(2) INFORMATION FOR SEQ ID NO: 281:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 901 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1311

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

CCGGCCACGG TTCCATCCAA CTTACAGGT GTGCACCTGA TTGTGTATGT AATTGTCAC	60
AACGGTAGAA TTTCACCTAT CCCTCCTATC TGCTCGCAGT ACCCGCAGAC TTTCTGAAAG	120
AAGAAGATAA CCTACTTATC CGTTGCTATG ATTATACTAA AGTTTCTACT TTTTGC AAA	180
TAGATTTTAA AATTTTGGC TAATTGTCTG AATCAGGGTC GGAAGTTTGA CGACCTGTG	240
ATTGCCTAGT TTTTCGGCTG CAATTTTGAG AATGGCACCT GAGTCTTTG AAGCAAAGAG	300
GAATTTTCCT TTGTCTGTAA AGACTTCGAA GTGGCGGCTG ATTTTGGCTC CAGTGACATT	360
GGCTCCAATC TGATTGATAT GGCTCCAAGG AATCTGGATA AATTGTTTGA CATTGACATC	420
TGGGTAAAAT TCCAAAGCCT GATCTCCGAC AAGGAATTTT CCAACTTTCC CAGCGATAGA	480
GAGGTAGGAA GTGCCTGTCT TACTGAGGAG TACTGTTTTG TTAAGTGATT GGGCCATGCT	540
TAGTCTTCCT TACTTTCTCC AAAAAAGGCA TTGTAGAGGG CTTTAATTGC TGCTTTCTCT	600
TGGTCTTTAT TGACAACAAA CATAATAGAA ACTTCACTAG AACCTTGAGA CATCATCTGG	660
ATGTTGATTT TGTTTTCAGA TAGAGCGCGT GTCGCAGTAG CAGTCACTCC GATATGGCTC	720
TTCATTTTTT CACCAACAAT CATAATGATA GAAAGGTCGT GTTCGATTTC TGCATGATCT	780
ACTTTAGCCT TTTGAACCAA CTGACGCAGG ATTTCTTCTT CTTGATGGG AGTTAGTTGG	840
CGAGAACGGA GAATGATAGA AAGAWCGTCG ATACCTGTTG GCATATGTTT CCAACCGATC	900
T	901

## (2) INFORMATION FOR SEQ ID NO: 282:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1765 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

CCCTGTTACG TGGATAATAG GGTAAGACTG CTCAGGATTT CCTAACAAAT CCACCGCTTG	60
CTGCATTCTGA CCCAAACCTG ATCGAAAATT CAAACCAATC CGACTATGGA GCCATTCTTC	120
TACTTCAAAC ATACACATCT CCTTGACAAA AGTCCAATCA ATTATCGCAT TAAAGTATGG	180
TTACTAATAA AAACAAGGCC AGGATTTTTCG TCCCGACCTC TTACCTGGTT AGCTAATAAC	240
TAGCTACTAT GAATGTGAAT ATGGGCTAAA AACATCCACT GGACGTTCCA ACTCTTCCCC	300

1312

ATTTCTGGGA GTTGGGGTAA AAATGTTTAC TGGACGTTCC AACTCTTCCC CATTTCTGGG	360
AGTTGGGCTG ATACAGTCTC CCAGACTGTA TCACTCCTCC ATAAAGCTGT TGAAGACTTC	420
TTCAATCATG TTCCATTCTG CTTCTGAGTC TTCTGGGATT GGTGCAATT CGCCTTCTGT	480
TCCATCTTCG TTTTCGATGA ATGAGTAAGC TTGGATTCA ACTTGTCCTG CTTCTGCTTC	540
TTCTGCGTTA ACTGGTACTA GAAGAACATA GTTTTTACCA AATTCTTCTT TTCCATCAAT	600
TGTCAAAAGG ATTTCAAACA AGGTTTCATT TCCTTGCTCA TCTACTAGTG TGATTAGTTC	660
ACGTTCTTCG TGGTCGTGGT TATGATCGTG TGACATAGCC TCGCCTTTAT ATTAAAAATT	720
TCTATCTAAA TAATTTTGTA AAATCAGCTG AGCTGCTAAC TTATCAATGA CTTTCTTGCG	780
CTTATTGCGA CTGATATCTG CTTGTTCAAT CAACATGCGC TCAGCAGCCA CTGTTGTCAA	840
GCGTTCATCC TGATAGTCTA CTGGTAAACC AAAAACTCT TCTAGCTTTG CTCCTGAGCT	900
TGACTAGCTT CTACGCGCGG TCCACTTGTA TTGTTTATGT TTTTAGGCAA GCCCACTACA	960
AATCGTTCCA CCTTGTAAGT ATCAACCAAT TCCTTAACGC GGTCAAAACC AAATTGGCCT	1020
TGTTCTTCAT TTATCTGGAT GATTTCAGC CTTTGAGCTG TAAAACCAAG CGGATCGCTA	1080
ATCGCCACCC CTACCGTTTT TGAACCGACG TCCAATCCCA TAATTCTCAT AGGTTATAGA	1140
TCGACTCCTT GTCCTTTGAG GTAGTAGCGA ACCAATTCCT CAACGATTTT ATCAGCTCA	1200
TACTTACGGA TTTGATTTTC TGCAATTATTA TAACGAGGAA CGTAGGCAGG GTCTCCACTC	1260
AATACGTAAC CTACGATTG GTTAATTGGG TTGTAaCCCT TATCGTTCAA CGAAGCATAA	1320
ACATCTGTCA AAGTTTCGCT AATTTCTTTT TTATTGGAAT CGTCCAATTT AAAACGTA	1380
GTTTCTTCAG TAAATCCCAT TCTAACACCC TCTTTCCTTA GAATAGTACC ATTATAGCAT	1440
AATTCCTTAC CTTCTACAAT TCAGGCAGTC TATTTATTG GATTTTCTAT TGTTCGTGCG	1500
CGCCATTTGC CAATCTATCT GAAATATATT TGCTTGGTTC ATTTTTCAAA AGATTTTCCA	1560
AACCAATATT CTTCAGATGT TCCAACGGG AAGCCTTCTT GACATCCAGA ACTTGAAAAT	1620
CAAACTAGT CGTTGTTTGA AGTTCCGTTG CGCTCAATAG TTTTGTTTCA AGTTTGAAAC	1680
CTGCCAATTT ACGAGCTTCA ATGATAGACT TATCCTTCTC CTCCGCTTCA AGAAGAGCTT	1740
TTTGAGTTTC CTCCACTCCA TGTTG	1765

(2) INFORMATION FOR SEQ ID NO: 283:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1346 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear



1313

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

CTTATCCATT CACTTTCTTG TCTGTTATTC TATAAATCTT ACTCCTAAGT ATACCACATT	60
TGCCCCCTAGA TGTGAACGAG AGAAACGCTC TAGACATTGC CAAGAAGGAA AAAAAAGGGT	120
ACAATGTAAC AAAATCAAGG GAGGTCTGGA ATGAAGAAAC AAAGCAAGTA CAAAGAGGTC	180
GTTTCCTATC TGA AAAATGG TATCGAGTCT GGACGATTTC CGACGGGTAG TCGCCTGCCT	240
TCTATCCGTC AACTGAGCCT TGACTTTTAC TGCAGCAAGG ACACCATTCA ACGAGCCCTG	300
CTGGAATTAC GGCACGAACA ATACCTCTAT GCCAAGCCTC AGAGTGGCTA CTATGTATTA	360
GAACAAGGGC AACATCAAGA CCTAGAAATC GAGGTTACCG ACGAACATGC CAGTGCCTAT	420
GACGATTTC GACTCTGTGT CAATGAAACC TTGATTGGCC GAGAAAACTA CCTCTTCAAC	480
TACTATGACA ATCAAGAAGG ATTAGAAGAC CTAAGACAGT CCATTCACAA ACTCCTCTTT	540
GAGCAAGCTC TCTACTGCAA GGCTAACCA CTAGTACTGA CTTCTGGAAC CCAACAAGCC	600
TTGTTTATCC TCTCTCAAAT ATCCTTTCTT AGACAAGCCA AGGAAATCTT GGTGGAACAG	660
CCAACCTACC ATCGGATGAA TCGCCTCTTG ATTGCACAGG GGCTGGACTA TCAAACGATT	720
GAACGAGGCA TTGATGGGAT TGACTTGAG GAGCTGGAAG GCCACTTCAA AACAGGAAAA	780
ATTAAGTTTT TCTACACCAT TCCCCGATTT CACTATCCCC TGGGACATTC CTATTCTGAG	840
CAAGACAAAC GATCTATTCT TAACTTAGCT GCCAAGTATG ATGTCTATAT CGTAGAGGAC	900
GATTATCTGG GTGATTGGA CTCCAAGAAG GGCCAAACCT TCCACTATCT TGATACAGAG	960
GAGCGTGTC TTTATATCAA GTCCTTCTCG ACCAGCCTTT TTCCTGCCCT TCGTATTACA	1020
GCACTCATT TCCAAATGC TATCAAAGAA GCATTTGTGG CCTACAAAAA TATCCTAGAC	1080
TACGACAGCA ACCTCATTAT GCAAAAGGCC CTGTCACTCT ATATTGACAG TCAATTGTTT	1140
GAAAAAATC GTTGGCTCG CTTGACCAAT CATGAATCTT ACCAAAAACA AATCGAGGAA	1200
AGGATAACTA AAACACCTTG TCCCCTTCTT CATTATTCCC TACACGATGG YTTATTGCTA	1260
GACCTGAGAC AGTATCCTAA AATCGCCAGT CTCAAACACA GTCAACTGGG CTTGGACTTC	1320
TTTGAAGAGG CCTATTTAAG CACCTG	1346

## (2) INFORMATION FOR SEQ ID NO: 284:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

1314

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

CTATATTCAG AATATGCCAA AAATTCGGAA TGGTATAAAT TTGCGGAGGG TTCATTTGAC	60
ATATTTAGAA AACTCCCCCA AAGAATTAAT TTAAAGAAAG ATTTTCTAG AATTTTGGCC	120
CCCTTTATTA TTAATTTGCT TAAATTAATC AATAATTATC TAGAGAATAA AGAATACGAG	180
TGGATTGACA AGAATGGAAA TATTTTTC TCTCTAGTAT TTTATTTAGA AGATTTAATC	240
TATCCTTGGA TTGTAAACC TTTGGTTTA GAGATAAAT CATTCGCTGA AAAAGGTTA	300
CTTGAAGGGG AATCGGAGCA GCAACGGTAC AAATATTTTA TAACATTGTT TGACAAGGAA	360
GAGAATATAT TAAATTTTA TAACAAATAT CCCGTTTAC TGAGGCAAAT ATCGGAGTCT	420
TGTCTTCGGT TCTATACTTA TTTTATAGAA ATTTTATCAA ATTTAGAAAA TGATTTAGT	480
GTGCTAGAAG AAGAATTAGG GCTAAGGGGG AAATTAAATG ATATAAAAT TGGAAAGGGT	540
GATACACACA GCCAAGGAAA AACTGTTTGG ATACTCTTCT TTGATGACGC GAAAATTGTT	600
TACAAGCCTA AAAATTTAAT AATCAATAAC TCACTAAATA CTATTGCTGA GTATATCCGA	660
AAGGTTGATG AAAAAATTAG GATAAGAATA CCTCGAACTA TTGCTTATTC GGATCACAGC	720
TATGAAGAAT TTATTGATTA TCTACCTCTA GAGCAAAAGA AAAATTTACC TGAATATTAT	780
TATAATTTTG GTGTGCTTTT AGCATTTATA TATTTATTTA ATGGGAGTGA TATACATTTT	840
GAAAATTTAA TTTCTATGG AGATATGCCT GTAATAATAG ACTTTGAAAC AATGTTACGG	900

(2) INFORMATION FOR SEQ ID NO: 285:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 862 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

TTATTTAGCA GAGGCAGTTT TAAATGTGAA GGATTGGTC AGTCAAACAG TTTTTATCA	60
GCAGATTATT GGTTTAGAAA TCCTATCTCA AACGGATACA GAGGTCGTTT TGGGACTTGG	120
AGGAAAAGCC TTGGTACACT TGATTCAAGC ACAAGAGGGT GGAGAAGTAA GGAACATTA	180
TGGTCTTTAC CATCTGGCTA TTCTTTTGCC GACACGAAAG GCTTTGGCGG ATGTCTTGAA	240
GCACCTGACG GATTACAGA TTCCTCTTGT TGGCGGTGCA GATCACGTTT ACAGTGAGGC	300
CCTTTACTTA GAGGACTTGG AGGGAAATGG CATTGAACTC TATCGAGATA AGCCAGTTTC	360
CACATGGGAT ATTCGAGAAG ATGGACGTAT TATCGGGGTG ACTGAAGTCC TTGCGGCTCA	420
GGATATCTAT GAGTTGGGGG AAAGAGTAGA GCCTTTTATC CTAGCAGAGG GTACGAGAAT	480

1315

GGGGCATATT CATCTTCTG TCAAGGATAG TCGAAAGTCC AGACAGTTTT ATCAAACGGT 540  
 GTTAGGGCTC GAGGATAAAT TCACTGTGCC TAGTGCTAGT TGGATCGCAG CTGGGGACTA 600  
 CCATCATCAT TTAGCAGTCA ACGAATGGGG AGGAAAAGGT CTGGATCCGC GTAAACAAGT 660  
 CCTACCAGGT TTAGCCTACT ATGTCATCGA AGTCGCACAT AAAGAAGAAC TGTTAACGAT 720  
 TGCCCAACGA GCACAAGAAG TTGACGCACC AATCAAATGG ATGACATCGA TCCAATTGGA 780  
 AATCACAGAC TCAGATGGCA TCGTGACCCG TATTCGTTTA GCTAGATAGA TGGTATGTGA 840  
 TGAAGGTAGA GCATCAATTG TA 862

(2) INFORMATION FOR SEQ ID NO: 286:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 650 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

TCGTTTACAA GATCGCTAAA ATGCATCTCA TGATCGCGAC CACGAATTCC AAGATAGCAC 60  
 GCGCTACCTC AATCATAGAT AGTTCACTTT TTTCTTGCCC AGCAAATACT TCTAATTCCA 120  
 AAGCGTTTCT CCTCATTTAT ACTACTATCG CCAGAGCGAA CAGACTCTGA CCTCATTTTA 180  
 TCATTTACTC TTTATTTTAC GATAATTTTG CGGAATAGTC AAAGGTTAAG GGGGAGAAAG 240  
 TGGCAGGATT AGACTAATTC CAATATAAAA CTCATTCCTT TTTCTGTTGC TCCATTTTCC 300  
 ACAAATCCAA GCGACTTGAA ACACCTCCTA GAAGCATGAT TGTAGGTGTA GATTTTCTTG 360  
 ACTCTCAATT CTTTCCATCC TTTTACTCGA GCCAATTCAA TCAAAGCACT TAGAATCTTT 420  
 TTTCCAAGTC CTCGATGTTG GTAAGCGGAA TTCCAATCA CAATGGGGAG ATTATCCTGA 480  
 GATAGTGTA TATCCCCAAT TGGAAACCAT TCTCCCTTCT CTTTGACTTC AATCCAAAAA 540  
 AGCTCACCAT GCCGATyCa ATAGGAATAC ATGGCTTCCA AGGTCGCTtG ACTGTAAGGA 600  
 AGCTTCACCC CATCTACGAG GtAAcCAAGT TCACATCCGT GATACCAAGC 650

(2) INFORMATION FOR SEQ ID NO: 287:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1119 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

1316

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

GATAGCAATC CGCTTCAGAA ACTTCTCGCT TACCTCTAAC TCCGATCGCT AGTTTGGGAG	60
AAGATACTTC CATTCTCATA CTATCTGTTG GCTTTCAGG CTGTAAAAAC AACTTTTCTC	120
TTGCTACTTC CTGAAAATCT GAATCTTGCA GTTCTTTGCT TTCAAAATAG TCCTGTACTC	180
GCTCCACATC AAAATTCCCA GCTAAAGACA GAGACATGTT TACAGGTTTG TAAAACTTTG	240
TAAAAATTTT TGCAAAATTA GTTAGATTGA TTTGGGAAAT GGACTCCTCA CTTCCAAC TA	300
TATCAGTTGC TAAAGGTGTA CCAGGATACA AATTCGCTAA AGTTGAAAAG AATAAACACG	360
AATCTGGATC ATCTTGGTAC ATTTCTCGTT CTGCTGAAT AATATCCTGC TCTGTCAGAA	420
TGGAAGCTTC AGTAAAGTGT GCTGATGTTA CCAATTCATC AAGTAAATCT AATTTTCTA	480
AAAAATAATC CGTTGCTGAA AAAAGATAGT TTGTTTTTGT AAAGCTTGTA AAGGCATTAC	540
TATCTGCACC TAGACTCGTA AAAGCCGACA TCAAACTACT AGAATCTTCT CTCTCAAATA	600
ATTTATGTTT AAGAAAATGA GCAATTCCTC CAGGATATTG TTTTACATCT CCGTCAACTT	660
CTGTGACAAA CGTATCTACC GAACCAAACT GTACAGTGAC ACTCCCGTAA ACCTCTTTAA	720
ATTCCTTTTT AGGCAAAAGA GCAACTGTCA ATCCGTTGGC CAAACGAGTT CGATAAACCA	780
TTTCTTTTAC AGCTGGATAG TATTTTTCTT CAAAACAAC CTTTGTCAAT CTATTCCTTC	840
CATAAAGTAA ATCGCTTGTA GTTTCACATT ATTAGCTACT CTACAAATAG CATCTTTGTC	900
AATTTGTTCA AGCTTTGCAA TCCAACTTT AAAGTCTGCT GAAGATTTTC CAAATAAGGC	960
ATTTTGATAA GCACGTTCAA TCAATGAAGA ATGATTATCT TGAGAAAGTA ACAACGACCA	1020
ACGAATCATT TCCTTGGTCT GATTAACTC AAACCTCTGTA AAAAAACCTT TTTTAAATC	1080
AAGCCGTTGA TTATTCATCA ATTTACGAGC CTGGTTACG	1119

(2) INFORMATION FOR SEQ ID NO: 288:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 540 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

ACGCCCTCGC GGGGACATGA CGAATTCCCC GTTCATCAGC AAGGCCGCCG AGGAGTGGGG	60
GGTGCCGTCC AAGTCAAAAG CGGCCCCACA TCGATTCAGT TCCCCGACGA ACAGCCCTTT	120
CCCCCAGCGT TCCTGGCTTT GCAACCGTTT CACAACAGCC TCGTAAAGTA GGCCGGACAA	180
GGCAGACGGA CTCCAAAGGA GTTCTTCCAT CTGCAAGTGC GCCTGCGTTA TGTGATCCCC	240

1317

GTCTTTTGCA TGTGTGTGGC ATGAATGCTG TTCCCAATCC CACTCCAGAA CATTCTCCTC	300
AAAAGTGCGC AACGTCGCCC TGAATGAATC CTGCCCTTGTA GTCGTGACCA TTCCTATGAA	360
GGGTGCGAGA GGATTTTCCC CGAGTGCAAG CGCATCCTCC GGCTCAAATC GGGTGCAATTT	420
CACAGTCCCG CTCAACGCTA GCGCGATCCC TTTTGGCAT GGTGACTCAA GCGTCCTTTC	480
AAACAAAAGC TCCTCATCCG CTCCAACCGG CCCGACGTAG ACGCGTAGAC CGAAGTCGTC	540

## (2) INFORMATION FOR SEQ ID NO: 289:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1949 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

AAAGAATTCG ACCAATTCAA GGTGAGGCA TCGCAAAC TAAGACTGTTT CCCCCTCAGT	60
TCTGGACAGA AAACGGGATA AGGTTGGCTG TGAAGCAAGC TGCCCTCCTA CCAACAATTT	120
TGGAAAGTAG GCATCAGCTG ACAATTCTTT ACAAGCATAG TCCGTTCCAT AACCTGTTAA	180
CAGTTGAAAG AGGAACTGGA CAAGGATATC TGAATCCGAA TAACGACAGT AGCGGCGTTG	240
GTCATTCGTT ACTAAATACT TAGAAATCCG CTCTTTTAGT TTCAACTGGG AAAAAAGTTC	300
CTGAAAAAAG ATAAGACCAC CATACTGGGT TAAATGACCT CCATCGAAAG ATAGTTGGTA	360
AAAAGACTTG TTTTGGAGT GATGATTGG TAACTGTTT ATGTGAGTTT CCTTCTTTT	420
TGTGTTTTTT TCTACACTTA TACCATAAAG GGGAACTCT TTTTGTCTA GTAAAAACA	480
CCCATTGGGT GAAAAAAGAA ACCATCCAGG ATCTAAGCTA AGGCAAGGAT TCTGGATGGT	540
TTTGTAGATT GGGGTGAATA ATTGGGTTT TACAATATCA ACTCCCATGA TAGTCATGAG	600
ATGACTCTTC ACGAATTGAC GTGATGACTG TCCTTCCTTT TGCATAATTA CCTCCGAAAC	660
ACAAAAAAG GGGTAGACAA TCTAGTGTCT ACCCCGAAA GTTTATTAAA AAAAAATCC	720
TGCCAAAGAA TTTTGGCAG GAAACCAAT CAATTTATCA GTTCTATCA ATCGCTTATC	780
GCTCTCAAAG ACTGGTAAAT AGGGATTCCG CAATCAAATT GCGATACTCT ATTATTTAAG	840
AGTAACTGAA GCTCCAGCTT CTTCGAATTT AGCTTTGATT TCTTCAGCTT CTGCAGTTGC	900
AACGCCTTCT TTAACAAGTG CTGGTGACC GTCAACAAGT TCTTTAGCTT CTTTAAGACC	960
AAGACCAGTG ATTTACGTA CAACTTGAT AACGCCAACT TTTTGTGCG CTGCAGATGT	1020
CAATTCAACG TCGAATGAAT CTTTAGCAGC ACCAGCATCA GCTGCATCAG CTGCAGCAAC	1080

1318

AGCTACAGGA GCAGCTGCAG TTACACCAA	TTCTTCTTCG ATAGCTTTTA CAAGGTCGTT	1140
CAATTCAAGG ATTGAAGCTT CTTTAATTT	AGCAATAATG TTTTCAATGT TCAATGCCAT	1200
TGTTATTTCC TCCAAATAAG TTTTAAATTT	TATAATAGTT TTTTTCGTAG CTAGkTACG	1260
CTGTGTAGCT TAAGATTAAG CCGCGTCTTC	TTTGCTTTCT GCAACCGCTT TGA	1320
AGCAACGTTG CGCACTGGCG CTTGAAGTAC	AGAAAGGAGC ATAGAAAGAA GTCCTTCGCG	1380
GTTTGGAAGA GTTGCAAGTG CAAGAATCTC	TTCTTTAGAT GCGACAGCGC CTTCGATTGC	1440
ACCACCTTTA ATTTCAAGTG CTTCAGCGTT	TTTAGAAAAG TCGTTCAAGA TTTTCGCTGG	1500
TGCGATAACA TCTTCATTAG AAAATGCTAC	TGCAGATGGT CCAACAAATA CAGATGCAAG	1560
ATCTTCAAGA CCAGCTTTTT CAGCTGCACG	ACGCAAGATT GAGTTTTTAA TAACTTTATA	1620
CTCAACTTCG CTCCACGAA GCTCAGGACG	AAGAACTGTA TCTTGCTCAA CTGTCAAACC	1680
ACGAGCGTCT ACAACGACGA TAGATGCAGC	AGCTTTCATT TTTTCAGCTA LACGTCAACT	1740
AGTTCCGCTT TTTTAGCAAT AATTGCTTCA	CTCATTAGTG TGTTACCTC CGTAATTATT	1800
TTGCTTGGGG AATTTTTCAA AAAGAAAAAC	GCGCCCAATC CTAGACACGA AAGTACAATA	1860
CGCTTCTTTT TACATGATAC GTTTTGTCT	CGGTAGGATA TTTATGAGTC GAGCTCCCT	1920
ACTGTCTTAG GCAGTTTTTT TAGATACGG		1949

## (2) INFORMATION FOR SEQ ID NO: 290:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1023 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

GGACTGTTTG ATCTTATACA GTAGCTGCTT	GATCCAAGCT TTCACCGATA GCGGCTAGGC	60
GCTCGATAAC TTCAGCTTGT GTCAATTCAT	TTTTTGAAAC ATAGCGGTTA CGTGGGTG.A	120
CACGGCACTC GTGTGAGCAT CCACGAAGGT	ACTTGTCTTC ATTTTCTTCT GATGTCAAGA	180
TACGACGGTT ACAGAAATGGA TTTCCACAGT	TGACATAACG TTCACATGGT GTTCCATCAA	240
ACCAGTCTTT CCTACGATA GTTGGGTGTA	CATGGTTGAC ATCAACGGCA ATACGCTCGT	300
CAAAGACGTA CATTTTCCCA TCCCAAAGCT	CACCTTGAAC TTCTGGGTCT TTACCGTAAG	360
TTGCGATTCC TCCGTGCAAT TGGCCGACAT	CTTTGTAGCC TTCACGGACC ATCCAGCCTG	420
AGAATTTCTC ACAGCGAACG CCACCTGTAC	AGTAAACCAC GACACGCTTG TCCATGAATT	480
TTTCCTTGTT ATCACGGACC CATTGTGGTA	ACTCACGGAA GTTGCGAATA TCTGGGCGAA	540

1319

TAGCTCCACG GAAATGTCCT AGGTCGTACT CATAATCGTT ACGTGTGTCA AGGACAACGG	600
TATCTTTATC AAGAAGCGCT TCTTTGAACT CTTTGGAGA CAAGTAAGCA CCTGTTGTTT	660
CAAGTGGGTT GATGTCATTG TCAAAGTCGT TGTCTTCCAA ACCAAGGTGG ACAATTTCTT	720
TCTTGTAGCG AACAAACATC TTCTGAAGG CTGTTCATT TTCTTCGTCA ATCTTGAACC	780
AGAGTTCTTC CATTCTGGA AGGCTGTGAA CGTAGTCCAT GTATTTTGA GTTGTTCAT	840
AGTCACCTGA AACTGTTCCG TTAATTCCT CGTCAGCGAC TAGGATACGG CCTTTAAGG	900
CGATTGATTT ACAGAAAGCC AAGTGGTCTG CAGCAAATTG CTCTGCATTT TCAATTGGAG	960
TATAAAGGTA GTAAAGTAAG ACACGAATAT CTTTGGkCaw AAGATTTGTA TCTCTTATC	1020
TAT	1023

(2) INFORMATION FOR SEQ ID NO: 291:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 3831 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

ACTATGAACA AGACCCAGAA AAAGTAGCCT TATTCTTAA GAATTTTAAT AGTTTAAAGC	60
ACCTAGCACC TGTTTAGATT GACGAAACAG GATTCGATAC TTATTTTAT CGAGAATATG	120
GTCGCTCATT AAAAGGTCAA TTAATAAGAG GCAAAGTATC TGGAAGAAGA TATCAGAGGA	180
TTTCTTTGGT TGCAGGTCTA ACAAATGGTG AATTAATCCG TCCAATGACT TACGAAGAGA	240
CGATGACGAG CGACTTTTTT GAAGCTTGGT TTCAGAATTT TCTCTTACCA ACATTAAACA	300
CACCATCGGT TATTATTATG GATAATGTAA GATTCATAG AATGGGGAAG CTAGAACTTT	360
TATGCCAAGA GTTTGGGCAT AAACCTTTAC CTCTTCTCC CTACTCGCCT GAGTACAATC	420
CTATTGAGAA AACATGGGCT CATATCAAAA AGCACCTCAA AAAGGTATTA CCAAGTTGCA	480
ATACCTTTTA CGAGGCTTTT TTATCCTGCT CTTGTTTCAA TTGACTATAT TAGAGGCGAG	540
ACATTTTTCG GTTCTTTGTC AACTGTAGTG GGTGAAGAA AGCGAAGATC TAGAAAGGAC	600
AAATTTTCGTC CTTTCTTTT TGAAGTTTTC AAAGTTTCTA AAACCAAAGG CATTGTGCTT	660
GATAAGTTTG ATGAGATTAT TGGTGGCTTC CAGTTTGGCG TTGGAATAAG GTAATTGAAG	720
GGCGTTGACG ATTTTCTCTT TATCTTTGAG GAAGGTTTAA AACAAAGTCT GAAACAGAGG	780
TGAAAAAGCA AGAGCTGATA GAGATTATAG TGGTGTTTAA AGTCTTCGGA ATAGCTCAAA	840

1320

AGTTTATCTA GAATTTCCTT ATTAGTCAAG TGCATACGAA AAGTAGGGCG ATAAAATCGT	900
TTATCACTCA GTTCTGACT ATCTTGTTGA ATGAGCTTCC AGTAGCGCTT GATAGCCTTG	960
TATTCATGGG ATTTCCGATG ATGGCTTGTG TTCTGCTCTC AAGAACAGTT ATGATATTGA	1020
GTTTATCAAA GTCCTGAGCA ATAAAGCTCA TCTCCATCTC CCGATTGAAA CAGTCACTCC	1080
CCGGACTGTT TCAACsTCCT AGGACATAAT CTCAGGAAGA ESEGAAAAAT CATGCTCAAA	1140
GTGAAAATCA TTGTTCTTGC GAATGACAGT TGAAGTTGAA ATAGACAACCT GATGATCAAT	1200
GTCGGTCATA GAAGTCTTTT TAATTAGCTT CTGAGCAATC TTTTGGTTGA TGATACAAGG	1260
AATTTGATGA TTCTTCTTGA CGATAGAAGT CTCAGCGAGC TCCATTTTTG AGCAATGATA	1320
GCACTTAAAA CGGCCTTTTC TAAGAAGAAT TCTAGTTTGA ATTTTTTTAT ACTAGAAAAAT	1380
CAGAACCATA ATACCTATAT AAAAATATTA TAGTCTAAT AGGATTTACC CAAAAGTTTT	1440
AAGGCGGTCT TTTTAGAACT TTAATTGTTT GAAATTTAGG TAGCAAATTT GTTCTATTT	1500
TGTCAACTTT TCCTATTTTT ATCTTGTTGA GGCTGGTATT TTAACAATTC AGGAATTGAT	1560
AGTGAATGTG TAAAATTTTT TGTTAGAATA AGTTTATAAA AAAGAAAACG AGTATTGAT	1620
TATGTTACAA AAAATTTATG AGCAGATGGC TAATTTCTAT GATAGTATTG AAGAAGAGTA	1680
TGGTCCTACA TTTGGTGATA ATTTTGACTG GGAACATGTT CATTTTAAAT TTTTAATTTA	1740
TTATTTAGTG AGATATGGCA TTGGTTGTCG TAAGGATTTT ATTGTTTACC ATTATCGTGT	1800
TGCTTATCGT TTGTATCTTG AAAAATTGGT AATGAATCGG GGTTTTATTT CTGTGTGAGG	1860
TAATTTTAGT AAATTTCCGA ACTAATTTAC TCTTTTATGG AAAGATGATA GTAAATAGCT	1920
AGTAATTTTT CTAAATCATT TTTTAATAGT TGGAAATAGC AAATCTTTCT ATTGTTTCTT	1980
CTTGATAAAA AGGCGATTTT TTATTATAAT AAATTGTAAG ATATAATTGC AGGTGAGAGT	2040
CCTGCCATGT ATGTGAGAAA GGAAGAGCCT GATGGCTCAG ACAAGATTAT GACTTCAGTT	2100
GTTGTTGTAG GTACCCAATG GGGTGATGAA GGTAAAGGGA AGATTACAGA CTTCTTTTCA	2160
GCGAATGCAG AAGTGATTGC ACGTTACCAA GGTGGTGATA ATGCTGGTCA CACGATTGTG	2220
ATTGACGGTA AGAAATTTAA GTTGCACTTG ATTCCATCTG GGATTTTCTT CCCTGAAAAA	2280
ATATCTGTCA TTGGGAATGG TATGGTTGTA AATCCTAAAT CTCTTGTAAG AGAGTTGAGC	2340
TATCTTCATG AGGAAGGTGT AACAACTGAT AACTTGCGTA TTTCTGATCG TGCGCATGTT	2400
ATTTTGCCTT ATCATATCGA GTTGATCGC TTGCAAGAAG AAGCTAAGGG CGACAATAAG	2460
ATTGGTACGA CAATTAAGGG AATTGGTCCA GCTTATATGG ACAAGGCTGC TCCTGTTGGA	2520
ATTCGTATTG CAGATCTTTT AGATAAAGAT ATTTTCCGTG AGCGTTTGA ACCTAACCTT	2580
GCTGAAAAGA ATCGTCTTTT TGAAAAATTG TATGACAGTA AAGCGATTGT TTTGATGAT	2640



1321

ATTTTGAAG AATATTACGA ATATGGTCAA CAAATCAAGA AATACGTGAT AGATACATCT	2700
GTTATCTTGA ATGATGCGCT TGATAATGGC AAACGTGTGC TTTTGAAGG TGCACAAGGT	2760
GTTATGCTAG ATATCGACCA AGGTACTTAT CCATTGTGTA CGTCATCAAA CCCTGTAGCT	2820
GGTGGTGTGA CAATTGGTTC TGGTGTCCGT CCAAGCAAGA TTGACAAGGT TGTAGGTGTA	2880
TGTAAAGCTT ATACGAGTCG TGTAGGAGAT GGTCTTTCC CAACTGAGTT GTTTGATGAA	2940
GTGGGAGAAC GTATCCGTGA AGTGGGTCAT GAATATGGTA CAACAACGG TCGTCCACGT	3000
CGTGTAGGTT GGTGACTC AGTTGTGATG CGTCATAGCC GTCGTGTTTC TGGTATTACT	3060
AACCTTTCTT TGAACCTCTAT TGATGTTTG AGCGGTTTG ATACTGTGAA AATCTGTGTG	3120
GCCTATGATC TTGACGGTCA ACGTATTGAC TACTATCCAG CTAGTCTTGA ACAATTGAAA	3180
CGTTGCAAGC CTATCTATGA AGAGTTGCCA GGTGGTCAG AAGATATTAC CGGAGTTCGC	3240
AATTTGGAAG ATCTTCCTGA GAATGCGCGT AACTATGTTT GTCGTGTGAG TGAATGGTT	3300
GGCGTTCGTA TTTCTACTTT CTCAGTAGGT CCTGGTCGTG AACAAACAAA TATTTAGAA	3360
AGTGTGTTGGT CCTAAGAGAT TTTTAAGATT TGTTTAAGAT AGGTCGGGTA TACTATAGAC	3420
GGTTACAAGA AGACCTCCTA ACTTGTGTA ACAAATATCC TAACTTTTC TTTTTCATAA	3480
TAATCTCCCT ATAGAGTCAC CGCATTCCGT GGCTTTTTTT GTGTTGGGAT TCATGATATA	3540
ATAATAAAAT CGATAAGTAG GAAAAGAGAA AAGAGATGTA TTATACGCTT GAAGAAAAAG	3600
AAGTCTTTAT GAGGGAGGCT TTGAGAGAGG CTGAGATTGC TCTTGAACAC GATGAAATTC	3660
CAATTGGTTG TGTGATTGTC AAAGATGGGG AAATCATTGG TCGTGGGCAT AATGCGCGTG	3720
AGGAATTACA GCGAGCGGTT ATGCATGCGG AAATTATGGC TATAGAGGAT GCGAACTTGA	3780
GTGAGGAGAG TGCGCTTGCT GGATTGCACA CTTTTGTGA CCATTGAACC G	3831

(2) INFORMATION FOR SEQ ID NO: 292:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1441 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

CCGCTGTTC AACCGCAACA TACCATAGTC CGTACGGGAT TCGAACCCGT GTTACCGCCG	60
TGAAAAGGCG GATGACTTAA CCCCTTGACC AACGGACCTG AGTTGTTATT TTCAACTCTT	120
ACTATTATAC AGTCTTTTCA AACTTTGTCA ACTACTTTTT CTAATTTTGT TTTATTTTTT	180

1322

CAACTTATAG TAAAAAAGC CAGAATTATA CTGACTCTTC TATCGCTCAT TAACTTAGA	240
AGCACGTTCT TTCCCCACC AATAAGGGAT TAGTTCTGCG ACTTTAACTG TTTTCTTAT	300
ATTATAGTCC ATCATGAATT CTGCATCTTT ATTTTCAGCA TTAAGCTCTA AAAGGAATTC	360
TCTACAAGCA CCGCAAGGCA TGGCTGAACT TCCACCATAA GGTGGTTTGT CTCGAAAGGC	420
TAATACTTTC TTAACCTTAG TTGTCCTGA AAATTGGTAC ATATTGAAGA GGGCCGCCCG	480
TTCTGCGCAG AGATGGA AAA CACCACAGGT TCCCTCCATA CAGAATCCTG TAAATATTTG	540
TCCATCTCCT GCTTCTACTG CAGCTACAAC ATGATTGGCA TAAACAAAGT CTGATACTTC	600
ATGTGGATTG TATAGTTTCT GTGCTTCTTC GTACATCTTT TCCCAGATGT CCATTATTGT	660
ATCCTCTTTA TTTAGAGATT TCTTTTAGCA TGTTTTCGAT ATGCTGAATT GATTTTTCAC	720
GTCCAAGCAA GAAAATTGTA TCTGGTAATT CTGGCCCATG CATTTGCGCT GAAACTGCCA	780
TACGAATAGG CATGAAAAGA TTTTCCCTT TAATACCTGT TTCTTTTGG ACTGCTTTAA	840
TTTGTGGGAA GATATTTTCT GTCACAAATT CATCATCTGT CATCGCTTCA AGTTTGTCTT	900
TGAATGCTTC AAGAACTGTT GGAAGTGTTC CACCCGTCAT GACTTCGCGC TCTGCTCTG	960
TCAATTCTGG GAAATCTGAG AAGAAAAGAT CTGTCAATGG GATAATCTCA TCTACTGATT	1020
TCAATTGTGG TTTATAGAGC TCAACTAATT TTTCAGCCTT GTCAGTCAAA CGGCCTGCTT	1080
CCTCTAAGAA TGGTTTGGC ATTTCAAAGA TGGTTTCAAG GTCTGCATTC TTGATATAAT	1140
CATTGCTCAT CCAGTCTAGT TTTTCTGAT CAAAGGCTGC TGGTGACTTG CTGAGGCGGT	1200
TTTCATCAAA AAGTTTAATG AATTCTTCAC GAGAGAAAAT CTCATCCCCA CCACCTGGGT	1260
TCCAACCAAG AAGAGCAATA AAGTTAAAGA CTGCTTCTGG AAGGTAACCT TTCTTTCGGT	1320
AATCTTCGAT AAATTGAAGT GTATTAGTAT CACGTTTAGA TAACTTCTTA CCAGTTTCAG	1380
AGTTGATAAT CAAGTGTCAT GTGACCGAAC TCTGGAGCTT CCTCAACCTA AGAGCGGGTA	1440
T	1441

## (2) INFORMATION FOR SEQ ID NO: 293:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4398 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

CGGCTTATGT AGTGGCAATC TTTCTACGTA AGCGAAACGA GGGGAGATTA GAGGCGCTAG	60
AAGAAAAAAA AGAAGAACTA TACAATCTTC CAGTAAATGA TGAAGTAGAA GCTGTAAAAA	120

1323

ATATGCACTT GATTGGACAA AGTCAAGTGG CTTTCCGTGA ATGGAATCAA AAATGGGTCG	180
ATTTATCTCT CAACTCTTTT GCCGATATTG AAAATAATCT CTTTGAAGCA GAAGGCTATA	240
ACCATTTCATT TCGTTTTCTC AAGGCCAGTC ATCAAATTGA CCAAATTGAG AGTCAAATTA	300
CTTTGATTGA AGAAGATATT GCGGCAATTC GCAATGCTTT GGCAGACTTA GAGAAGCAAG	360
AATCTAAAAA TAGTGGTCGT GTTCTTCATG CTTTGGATT ATTGAGGAA CTTCAGCATA	420
GAGTTGCTGA AAATTCAGAA CAGTATGGTC AAGCCTTGGA TGAAATTGAA AAACAATTAG	480
AAAATATCCA ATCTGAATTT TCACAATTTG TAACCTTGAA TTCATCGGT GACCCTGTGG	540
AAGCCGCAGT GATTTTGGAT AATACAGAAA ATCACATTTT GGCCTTAAGT CATATTGTGG	600
ATCGTGTTC AGCCTTGGTT ACGACGCTTT CTACAGAATT GCCAGATCAA TTACAGGATT	660
TGGAAGCCGG TTATCGTAAA CTAATTGATG CTAATTATCA TTTTGTGAA ACGGATATTG	720
AAGCGCGTTT CCACCTGCTT TATGAAGCAT TCAAGAAAAA CCAAGAGAAT ATTCGTCACT	780
TGGAATTGGA TAATGCCGAA TATGAGAATG GACAGGCACA AGAGGAAATC AATGCCTTGT	840
ATGATATTTT TACTCGAGAA ATTGCTGCTC AGAAAGTAGT GGAAATCTA CTTGCAACTC	900
TTCCAACCTA TCTTCAACAT ATGAAAGAGA ATAATACTTT ATTGGGAGAA GATATTGCAC	960
GTTTGAACAA GACCTATTTA CTTCTGAGA CAGCTGCAAG CCATGTTCTG CGTATTTCAGA	1020
CAGAATTAGA GAGTTTGGAG GCAGCTATTG TTGAGGTAAC TTCAAATCAA GAAGAACCAA	1080
CCCAAGCTTA TTCAGTTCTT GAAGAAAATC TTGAGGATTT ACAAACTCAA CTAAGAGATA	1140
TTGAAGATGA GCAAATTTCA GTTAGTGAGC GCCTGACACA AATTGAGAAA GATGATATTA	1200
ATGCACGTCA AAAGGCCAAT GTTTATGTCA ATCGTCTCCA TACTATCAAG CGATACATGG	1260
AAAAACGCAA TCTGCCAGGT ATTCCACAAA CTTTCTTGAA GTTATTCTTT ACGGCAAGCA	1320
ATAATACCGA GGATTTAATG GTTGAGTTAG AACAAAAAAT GATTAACATT GAATCTGTTA	1380
CCCGAGTTCT TGAAATTGCA ACGAATGATA TGGAAGCTTT AGAAACGGAA ACTTATAATA	1440
TTGTACAATA TGCAACTTTG ACAGAGCAAC TCTTGCAATA TTCTAACCGC TATCGCTCAT	1500
TTGATGAACG CATTCAAGAA GCATTTAACG AAGCTTTAGA TATTTTGTAA AAAGAATTTG	1560
ATTATCACGC TTCATTTGAC AAGATTTCTC AAGCATTGGA AGTGGCAGAG CTTGGTGTA	1620
CCAATCGCTT TGTTACCTCA TATGAGAAAA CACGTGAAAC GATTCGTTTT TAATAAAGA	1680
AAAAGATTTT ATTGTGTGAG GAGCAGAATC AAATCTTTT CTATAGTTGT GGGGAGATTT	1740
ACTTCATTTT CTCCTGAGAT TGAGTTTTTG CCCAGCCGAT TTATCCACTA CCTCAAAACA	1800
GTGTTTTATA CTCTCGAAA ATCTTTTCAA ATCAGTCAG CGTCGCCTTA CCGTACTCAA	1860

1324

GTACAGCCTG AGGCTAGCTT CTTAGTTTGC TTTTGGATTT TCATTTAGTA TTAAAGTGAT	1920
TTCCGCCAGTC TTATCTGCAG CTTCAAATCT GTACTTTGAG TAACTTGGTA ACCGTCCAAT	1980
AACGAAGTCT ATTGAAAAAT CTCCAGACTA GAGAACTCAC GGATAGTTCC TAATCTGGAG	2040
ATTTCTTATT TGCACITTTT TGTACAACCT TTAGTCCACG GTAAATAGAC CTCTAAAACC	2100
TCTTTGTTTA CGAGAGTTTC CTCGTTTGGG AGACATTCTA GAAGATAGGA TAGATATTTT	2160
TCGCTATTTA TACTAGACTA AAATCAAAAA GCATTATATA ATAGTGATAT GAAATCAACT	2220
AAAGAAGAAA TCCAAACCAT CAAAACACTT TTAAGAGACT CTCGTACAGC TAAATATCAT	2280
AAACGCCTTC AAATCGTTCT ATAGTAAAT GAAATAAGAA CAGTACAAAT CGATCAGGAC	2340
AGTCAAATTG ATTTCTAACA ATGTTTTAGA AGTAGAGGTG TACTATTCTA GTTCAATCT	2400
ATTATATTTT GTCTGATGGG CAAATCTTAT AAAGAGATTA TAGAACTTTT ATAGTAGATT	2460
GAAATAAGAT GTGAACAACCT CTATCAGGAA AGTCAAATTA ATTTATAGAA ATATTTTAGC	2520
AGCCAAGGTG TACTGTTATA GATTCAATAC ACTATAGACT GTAATCAAAC AACGATTGG	2580
CGAAATGTAA AAAAATATGA GGAGTTCGGA CTCGACTCTC TCCTTCAAGA AACACGTGGT	2640
GGTCGTAACC ATGCATATAT GACAGTTGAG GAAAAGAAAG TCTTCTTGC CCGCCATTTG	2700
AAGGCTGCAG AGGCAGGAGA ATTTGTTACA ATTGATGCCT TATTTCAAGC TTATAAAAAG	2760
GAGTTAGGTC GTTCCTACAC ACGTGATGCC TTCTATCAAC TGTGAAGTG CCATGGTTGG	2820
CGAAATATTA TGCCACGTCC AGAACATCCT AAGAAAGCAG ACGCTCAAAC CATTTGTCGG	2880
TCTAAAAATA AAATCTCAAT TCAAGAAGAA AAGAAAGCGC TTTAAAACCA GTAGACGTTT	2940
TCGTAAGGTT CGCTTGATGT ACCAAGATGA GGCTGGTTTC GGTAGAATCA GTAACTGGG	3000
ATCTTGTTGG GCTCCAATAG GAGTAGGTCC ACATATCCAT AGTCACTATA TACGAGAATT	3060
TCGCTATTGT TATGGAGCTG TTGATGCCCA TACAGGCGAA TCATTTTCT TAATAGCTGG	3120
TAGATGTAAT ACTGAGTGA TGAACGCCTT TTTAGAAGAG CTTTCAACAG CTTATCCAGA	3180
TGATTATCTT TTAATCGTTA TGGACAATGC TATATGGCAT AAATCAAGTA CCTTAAAGAT	3240
TCCGACTAAT ATTGGTTTTA CCTTTATTCC TCCATACACA CCAGAGATGA ACCCCATTGA	3300
ACAAGTGTGG AAAGAGATTC GTAAACGTGG ATTTAAGAAT AAAGCCTTTC AAACCTTTGA	3360
AGATGTCATG AATCAACTCC AAGATGTTAT ACAAGGATTG GAGAAGGAGG TGATAAAGTC	3420
CATCGTTAAT CGGAGATGGA CTAGAATGCT TTTGAAAAC AGATGAGTAT AAAAAGAAAG	3480
TCCTCATTTC AATAGAAATC ACGACTTTCT GATGGATTTA TAGTAAATG AAATAAGAAC	3540
AGGACAAATC GATCAGGACA GTCAAATCGA TTTCTAACAA TGTTTTAGAA GCAGAGGTGT	3600
ACTATTCTAG TTTCAATCTA CTATATTTTT GGAGTGATAG AAAAGCCCTT CATAAGCTAG	3660

1325

TCTACTTGTT CAGGTGCGAG AGCTTTGACA TCTTTTCTG TACTTAGCCA AGTCAGTTTT	3720
CCGTTCTCAA AGCGTTTATA TAGTAGCCAA AATCCTTGAC CATCCCAGTA AAGGGCTTTA	3780
AAGCGGTCTT TACGTCCACC ACAAAGAGA AAGACTTGAC CGGAGAAAGA ATCCAATTCA	3840
AAGTGGGTTT TAACTACATA GGCTAATGAG TCTATTCCCT GCCTCATATC TGTCTTGCCA	3900
CAAACAAGGT GAACTTGACC TAAATCACTT AGTTGAATTA TCATAGTACA ATACCTTTCC	3960
TCCGATAATT ATTTTATATC TAGTATACTG GAAGTTGGGG AATTAGGATA GATACCTTGT	4020
TATGACGCGC TTACGTAAC TGTAACTAGC TGCCTAGTTT GATCTTTGCT TCTTCATTGA	4080
TTAGCAGTAG ATTTCAAAAT GATAAAAACG CATAGTATCA GGTATTGAAA TGTACTGCCC	4140
CAAAAGTTAG ACAGAAAAAA TCTAACTTTT GGGGTGTTTT TGTTATGAAA TTAAGTTATG	4200
ATGATAAAGT TCAGATCTAT GAACTTAGAA AACAAGGATA TAGCTTAGAG AAGCTTTCAA	4260
ATAAATTTGG GATAAATAAT TCTAATCTTA GGTATATGAT TAAATTGATT GATCGTTACG	4320
GAATAGAGTT CGTCAAAAAA GGAAAAATC GTTACTATTT TCCTGATTTA AAACAAGAAA	4380
TGATTAATAA AGTCTTAC	4398

(2) INFORMATION FOR SEQ ID NO: 294:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 718 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

AGATTTTATAG ACTTTGTCTT TAATCGTTTC TTTTAGGGA TGATTGCGAC ACCTTCTTTT	60
GGCTATTAAC TTTAGCAGGA GGGATTATCC TTGGTCTAGC GCCGGCTAGT GCCACCTTGA	120
TGAGCTTATA TGCAGAACAT GGTATAGCT TTCGGGAATA CAGTTTGAAG GAGGCTTGGT	180
CTCTTTACAA GCAAAATTTT GTCTCAAGCA ACCTGATTTT CTATAGCTTT TTAGGTGTGG	240
GTCTAGTTT GACCTATGGT TTGTATCTCT TGGTGCAATT GCCTCATCAG ACCATTGTTT	300
ATTGATTGC GACCCTTTGG AATGTCCTAG TAGTTGCCCT GATCTTTTGG GCTTATACAG	360
TATCTTTAAA ATTACAAGTT TATTTGCCT TGTCTATCG AAATAGTCTC AAATTATCCT	420
TGATTGGCAT CTTTATGAGT CTAGCAGCTG TGGCTAAGGT TCTCCTTGGG ACTGTGCTAC	480
TTGTAGCAAT TGGTTATTAT ATGCTGCCC TGCTATTTT TGTAGGAATT GGGATGTGGC	540
ATTTCTTTAT CAGTGATATG TTGGAACCTG TCTATGAAAT CATCCATGAA AAATTGGCGT	600

1326

CAAAATAGAA TGAAGCAGTT TTGGCTACAT ACGCTTCTAA GAACCTATAG TTCAGTGATG 660  
 ATCATTATCA TTGCGAGTTT TGCAATCTTA CTCTCTTACG CTGTCTGGGA TTCACGTG 718

(2) INFORMATION FOR SEQ ID NO: 295:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 718 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

TCGGTACCAA AATTCTGGAT TTATACTAGC AAAGATCCAA GAGCAAATTA TTAAACAGAT 60  
 TTAGGTCTAG TTTCCCTGA ATCATTAAAA GAATTTGAGA GTGAAGATAG TTTTGCAAAG 120  
 GAAATTTCTG CAGAAGAAGC AAATAAGATA AATGATGCTG ATGTAATCAT AACTTATGGT 180  
 GATGATAAAA CTCTTGAAGC TTTACAAAAA GATCCTCTTT TAGGTAAAT AAATGCAATT 240  
 AAAAATGGTG CCGTTGCTGT AATTCCAGAT AATACACCGT TAGCAGCCTC ATGCACTCCA 300  
 ACACCACTTT CAATAAACTA TACTATTGAA GAATACCTAA ATCTTTTAGG AAATGCATGC 360  
 AAAAATGCGA AATAAAAAAC AAATAAACCT AGGCATAATT TTTATAATCT GCCTAGGTCT 420  
 TCTTATTACA ATATTTTGT CATTAAAGCT TGGAACAAAA GAAATTAATA TCAGAGATTT 480  
 TTTAGCAGCT TTTGGAATGG GTAATACAAA TGATGATTTT ATTAAATCAA TTATATATAA 540  
 TGAATACCT AGAACTATTT TTGCAATTTT AGCAGGTTCT AGTCTTGCCA TAAGCGGTGT 600  
 ATTGATGCAA TCAGTTACTA GAAACCCAAT AGCTGATCCA GGTATACTCG GTATAAACAC 660  
 AGGAGCAAGT CTTAGTGTAG TAATTGGTCC TTCCTTTTAG GGAATTCATC AAGCATAA 718

(2) INFORMATION FOR SEQ ID NO: 296:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1436 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

GAACTAATCA TTTTACAGG ATGAGATTTA CAGCAGAGAG TTTGAAGGCT TTATCAAAGG 60  
 TTTTCTTGG CATAATGACT TTTCTCGTT TCCACTTAAT TTTGTGTCTA CTTTATTATA 120  
 CCAAGTCCAC SCTTAAGTTA GATAATAAAT CTAACCTAAG GAAGCTAGAA GGATGAGAAT 180  
 CCAGGTGGTC AAGAGTCCCA AACTTAAGCT GATGGGGACA CCCAGAATAA TTTGCTTTTT 240

1327

GAAGGCAAGG CCACGTTCTT CTATATTGGG AAGTGAGAGT TGAATGAGAG AACCAGCTGA	300
TGAAAAGGGT GAGATATTAG TAGATAGAGC GCCAATAACG GTGGCTGTTG TGACTAAGTG	360
AATATCAATC TGAGGATTTT GAGCACTGAT GATAGCAATG ATGGGAAAGA GGGCTGGAGC	420
TACAACGGAT AGGGTGGAA TAAAGAGTGA CATCACTCCG GCTATCACAC AAAAGAACAG	480
AGGTAACCAG AAATGAGGAA TGGTTGTTGT CATGAGGTGC CCTATCAGTG TACTAAACC	540
TGACTTGACC GCTAGAGACA TTAGTAAGCT CATGCCGAG AGCATGATAA TTGTAGCCCA	600
GGGAACCTTA GCTAAATGG CTTCTTGCTT CCCTAATTTG AGCCTTAAGG CGAGGCAGAC	660
CATGAGTATT GAGACAAAGC CAATATCAAA TGTTTTTTGA TAAGTAGCTA TCCAGGCGAT	720
GTTTGGGAAA ATGAGATGCA ACAAGGGAAA AAGCCAAACC AAAACCATGC TGCTGATCAT	780
GAGCAAGGTG GTTTGTCTTT GAACCTTGCT GAGGAGTGGT GGTGGTCAA TAGTCAAGGA	840
TGAGTTTGT CTTCCCTTAC TATAGTGAAT GTAACAGGAT AATAAAGCA AGACGATGAG	900
TGGGTAGATA ATGCTGACGA TAAAGATATG ATTGCCAAGT GAAAAGCTT GCTCTTCCCA	960
TCCCATTTGC TTAAACAGGC CTTGAAAGAC AATGCCTGAG CTAAGGTGTA TCAAATTAGC	1020
CCCTCCTGAA GCTCCCAAT TGACGGCTTG AGCTCCAATC AAAGGGTGTG TGTCCGCTTT	1080
TTGACAGAGG GTAATCGCTA GAGGACAGCA AACGGCCATA GTAGTAAAA ATCCAGCACC	1140
TAAAGCAGAC AAAAGGGTTG CCATCAGGTA TAAATCATG TAGAGGGCGT TAGGGTGGGT	1200
GCGTGTGCGG TAGAGAATGT GTTGAGCCAA AACATCAAGA GTACCGTTAG TTGTTGCAAC	1260
GTTATAAAG AGAGAGACGC TAAAAATGGT AAAAAAGAGT GAGGTTGGCC AAAAATGAAG	1320
AAGTTCTTTG GGGCTTAATC CCATGAGAGT GGTGCGATG AGGTAAGAAA AAGCAATAGC	1380
CAGCAGGCCA ATATTGATTT TGGTGCGGTA ACCAATTCCA ATGGCTAGAG CAATGG	1436

(2) INFORMATION FOR SEQ ID NO: 297:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1696 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

CCATTGGGA AAGAACGTAA GAGTTTGCAG GGTGAGATTC CAGAAGAATT TTCAATGTCA	60
GCCGTTGACA TGTCTATGAT TGACCACATT CCAGATATGA TTGAAAATGG TGTGGACAGT	120
CTAAAAATCG AAGGACGTAT GAAGTCTATT CACTACGTAT CAACAGTAAC CAACTGCTAC	180

1328

AAGGCGGCTG TGGATGCCTA TCTTGAAAGT CCTGAAAAGT TTGAAGCTAT CAAACAAGAC	240
TTGGTGGACG AGATGTGGAA GGTGCCCCAA CGTGAAGTGG CTACAGGATT TTACTATGGT	300
ACACCATCTG AAAATGAGCA GTTGTTTGGT GCTCGCCGTA AAATTCCTGA GTACAAGTTT	360
GTCGCTGAAG TGGTTTCTTA TGATGATGCG GCACAAACAG CAACAATTCG TCAACGAAAT	420
GTCATTAACG AAGGGGACCA AGTTGAGTTT TATGGTCCAG GTTTCGGTCA TTTTGAAACC	480
TATATTGAAG ATTTGCATGA TGCCAAAGGC AATAAAATCG ACCGCGCTCC AAATCCAATG	540
GAACTATTGA CTATTAAGGT GCCTCAACCC GTTCAATCAG GAGATATGGT TCGTGCAATTA	600
AAAGAAGGAC TCATCAATCT TTATAAGGAA GATGGAACCA GCGTCACAGT TCGAGCTTAA	660
GAAAGGAAAA GGAAATGATA GAGGCACAGG GTTCTTAGT GGATAAGCAA ACAAGATGCA	720
TTCATTACCA TAGCAAGCTG GATATTATG CTTTACAATG CTATGATTGT AAAAAGTATT	780
ATGCTTGTTA TCGGTGTCAT GATTCATTAG AACATCACCC TTTTGAGCCG TATCCCTTAT	840
CTTTGATACA GGATAAGCCT ATTTTATGTG GTGTTTGTCT AAAACTACTA ACATATAAGC	900
AATATAAAGA AAGCTTAAGT TGCCCCTTT GTTTTCTCG CTTTAATCCA GGTTGCCAAA	960
ATCATAAGGA ACGCTATTTT AAATAGCAAA TCATCTAGTT TTGAAGTAGG AGAAAACTCA	1020
ATTTCAAGAG AAAATGAAGT AAATCTTCCC ACAATAAAC GCATAATATC AAGATTGTTT	1080
AATACCTGAT ACTATGCGTT TTTAAGATTT TAAAGACTTT TTTCTTTAT CTGGTATTTT	1140
GACTACTTGT TAAAACTGGG TTAATTTTCG ACTGTTTAAT AGTTATTATG CAAAGTCTAA	1200
AAGGTTAGAA TTGTCAAAAC AATCCGTCTA GAGTATGCGT GATGCCAACC GTGGTGGATG	1260
TTCTCAGTCA TGCCGTTGGA AGTACGACCT TTACGATATG CCATTTGGGA AAGAACGTAA	1320
GAGTTTGCAG GGTGAGATTC CAGAAGAATT TTCAATGTCA GCCGTTGATA TGTCTATGAT	1380
TGACCATATC TCAGATATGA TTGAAAATGG TGTGGACAGT CTAAAAATCG AAGGACGTAT	1440
GGAGTCTATT CACTATGTAT CAACAGTAAC CAACTGCTAC AAGGCGGCTG TGGATGCCTA	1500
TCTTGAAAGT CCTGAAAAGT TTGAAGCTAT CAAACAAGAC TTGGTGGACG AGATGTGGAA	1560
GGTTGCCCAA CGTGAAGTGG CTACAGGATT TTACTATGGT ACACCATCTG AAAATGAGCA	1620
GTGTTTGGT GCTCGTCGTA AAATCCCTGA GTACAAGTTT GTCGCTGAAG TGGTTTCTTA	1680
TGATGATGCG GCGGTA	1696

(2) INFORMATION FOR SEQ ID NO: 298:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1022 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear



1329

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

```

CCGAGTTTAT TATGGTTTCT TCGGAATTTA TCTCAAAGAT TGAATTTGCT TGCAATAAGA      60
AAGAAAGTCT TTATAGTCAA AGCAAATTTA AGTATGCGAT TCGTTCGATG TTCGCAGGTG      120
CATTTTTAAC CTTCACTACT GCTGCAGGTG CAGTTGGGGC TGACTTGATT AATAAAATTG      180
CACCAGGTAG TGGACGCTTC CTCTTTCCAT TCGTTTTTGC TTGGGGCTTG GCCTACATTG      240
TTTTTTTGAA TGCCGAGTTG GTCACCTCAA ACATGATGTT CTTGACTGCT GGTAGTTTCT      300
TAAAAAAAT CTCTTGAGA AAAACAGCTG AGATTTTACT ATACTGTACC TTGTCAACC      360
TTATCGGAGC CTTGATAGCA GGTGCGGCT TTGCTCATTG GGCAGCCTAT GCGAATCTGA      420
CACACGATAG TTTCATCTCA GGTGTTGTTG AGATGAAGTT AGGCCGCTCA AATGAATTGG      480
TCTTGCTTGA GCGGATTTTG GCAAATATTT TTGTAAATAT TCGGATTCTG TCATTTATTT      540
TGGTCAAAGA TGGTGGTGCC AAACCTTGGC TTGTGTTGTC AGCTATTAC ATGTTTGTAT      600
TCTTAACAAA CGAGCACATT GCGCGGAACT TTGCTTCTTT CCGGATTGTG AAATTCAGTG      660
TTGCTGCGGA TTCAATTGCC AACTTCGGTG TTGGAAATAT GCTTCGCCAC TGGGGTGTGA      720
CTTTCATCGG AAACCTTATC GGAGGAGGCC TCTTGATGGG TCTTCCATAT GCCTTCCTCA      780
ATAAAACGA AGATACTTAT GTAGATTAAG AAAATGAGCA CGATTGAGTC GTGCTTTTTT      840
CATTTTCAAA ATAAGGTAAT AGCTATTTCT TATATCAAAA TATAGAAAAC TGATATTTGT      900
A-ACTATAAC TCAAGGTGCT ACAATATCCT TAATAAAATA ATATGGAGGT CACCTTATGA      960
CTTG TGATTT TAAATTTGAA ACTCTACAAC TACATGCTGG TCAAGTTGTG GCTCCAGCTA      1020
CT                                                                                   1022

```

## (2) INFORMATION FOR SEQ ID NO: 299:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 663 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

```

CCTTAAGTAA TCTCTGATAA TATTTTCTTT ATTAGCATAG GGGAATATCG ATATAATGGC      60
TTCATTATGA GTGGCAGGAA TATCCAATAT GGCAACTTTT CCAATAGATA ATTTAAACT      120
CATTAAATAA GTTCCTTTAG GTGAAATGTC TATTTTCTTT GATTTTAATG CTAATTTAGA      180

```

1330

AATAGATTCT CTCGCATTAG TTACATAACC AGATATAGGC ATATCTGATA TAGATACCCA	240
AGGTATTTCA GTTCCCCAAA AAGTAGCTTC ACTGCGTGGA GGAGTTTTTC CTATTCTGAA	300
GTTAAGTAGG CTAGCAAATT TAATATATCT CCATGCTTCT GGGATTTCAT ATATAGGATA	360
AGAGGTGTT TCGTCTTTGT TCCATAATA AGAGTTATCA TCTCCTTGGG AAACAATAGA	420
AATGTCCAAA TCTTTCTTTT TAATCTTGCC TTCTTCAAAG AGTTTTTGT TTTCTGCTCG	480
TATTTTTTCA AGTAAACTT CGACTGATTC ATCATTTGGG TCTTGTTCAA CTAATTTTCC	540
TTGCATAGCA TATTGAAGAA TAGATTTTTT TAGTTTATCT GGAAATTCTT TATCTAGCTG	600
TTCTAGTCTA TTATAACTTT CAGCATATTC ATCTACTTTT TCTAAAGCTG ATTCGATTGC	660
TTC	663

## (2) INFORMATION FOR SEQ ID NO: 300:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 881 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

CGTCGCTGAA CATGTCAACA GCAAAATTAAA CTAAACAAAC TAAAATTATG TGATACTTCA	60
CATAATTTTC TTTAGAAAAT ATTATCAGAA GAAAGTTGAG AAAAATGGCA GAAAAACAT	120
ATCCTATGAC CCTTGAGGAA AAGGAAAAAC TTGAAAAAGA ATTAGAAGAA TTGAAATTGG	180
TTCTGCGACC AGAAGTGGTA GAACGCATTA AGATTGCCCG TTCATACGGT GACCTTTCAG	240
AAAACAGTGA GTACGAAGCA GCTAAGGATG AACAAGCCTT TGTCGAAGGA CAAATCTCTA	300
GCTTAGAAAC AAAAATCCGC TATGCTGAAA TCGTCAATAG CGACGCAGTT GCCCAGGACG	360
AAGTAGCGAT TGGTAAAACA GTCACCATCC AAGAAATTGG TGAGGACGAA GAAGAAGTTT	420
ATATTATCGT AGGTTTCAGCT GGTGCAGATG CCTTTGTAGG TAAGGTTTCA AATGAAAGCC	480
CAATTGGGCA GGCCTTGATT GGCAAGAAAA CAGGTGATAC AGCAACCATT GAAACGCCTG	540
TTGGTAGCTA TGATGTAAAA ATCTTGAAGG TTGAAAAAC AGCCTAAAAA CAGAAAAAGG	600
AGTGGGGAGG CGATGTGCTT CACTCACTCC TTTTTCATT TTGCTACTCT TCGAAAATCT	660
CTTCAAACCA CGTCAGCGTC GCCTTGCCGT ATGTATGGTT ACTGACTTTG TCAGTTTCAT	720
CTACAACCTC AAAACAGTGT TTGAGCTAA CTTGTCAGT TTCATCTACA ACCTCAAAAC	780
TATGTTTTGA GCTGACTTCG TCAGTTTCAT CTACAACCTC AAAACCATGT TTTGAGCCGA	840
CTTCGTCAGT TTCATCTACA ACCTCAAAAC TATGTTTTGA G	881

1331

## (2) INFORMATION FOR SEQ ID NO: 301:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 949 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

```

CCTTTTAA TACAAGTTAT TTTGATTAA CCGGCTTGTG TTAGCTGTG TGCAAAGCTG      60
TGGCAATCGT ATCTGCATAC AATTTTGCTC CTGCTTCGAT AGTGCTACTC TCACTCCCGA    120
AATGAACCTG GTCTGTTCCA GCCCAAATTT CTGGATGCTC TTTCGCAACT TGATTCCAAT    180
CTGCTATCGT AATGTAAGGT GTCTTCTCTG CCAATTCTCT CATATAGGCA GCAGCCTTCT    240
CAACGATGGC ATAGGTCTCT TTTGTCTTAT CTCCCTCATA AGGAGTCACC AAAATCATAT    300
GGTGTCCCTT AGGAAGATTT TTCACGATAC TGTCCTCAGTC ATCCTTGTA TTTCTCAGGAT    360
TATTTACCCC AGTCGCAATG ACCACCGTCT TAGGTAAAAA TTTATTCTGG CTATTATTTA    420
GCATGATTTT ATTTGCGGTC TTGGTTGTTA CGCTGACCTG CGCGTTAATC TGTGCTCCAG    480
GAAGAGCTGT CTGTAGTGCT GTATTTGCCC TTAAGGCCAC TGAGTCACCA ATTAACATAG    540
TGCCATCAGC AATTCCTAAA CTGTTTGTCAT CTGCCCGTTC TGCCATCACC TTGGTCTGGC    600
CAATATTTGT TGCAGCTTGC TTCAAGCCAT TGACAGTCAA GTCTGTCTCA AACGCTCCCA    660
CTTGTGGTGC CAACAAGGTC ACCGTGCAGA CAATGATGGT CAAGATTCCT GTACCTGCTG    720
CAAGAATTGC GTGAATATAA GGCAGGGGAC GAAsGGTTTG GACAATAGGT GTGTTCTTGC    780
CTGCAATCCA AGGTTCCAAT ACATAAAATG ACAGACTGGC AAAGCCATAA GAACAAATCA    840
GAGTCAGTAA TACAGCAAGA AGATTTGATG TCAACTGTGA GAAAATGATA TAGAAAGGCC    900
AATGGAAAAG ATAAACCGCA TAGCTAGTAT CCGCTAAAAA GCTGATAAT                949

```

## (2) INFORMATION FOR SEQ ID NO: 302:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 622 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

```

AAGATATATT TTTTACACAG AAGTATGCAA AAGTAAAGAG TGCAAAAAAT GGAATTAAAG      60

```

1332

CGAAAATAAA AGCCGTGTAC AGGCGACCAA ACCAACGTAC ACGGCTAAGG AAAAATAACA	120
AAACTCAAGC AAAGGCAAGG CGCGTGGTTT TGTAGGTAT TTAGCAAGGG GACAAACCCC	180
TTTGTAATA ATCTCCTCTT ATTTTATCAA AATTAGAGGA AAATGACAAC TTAATTTATA	240
AAAAGGAAAA ATGGAGGATA TAAATGGAAA TTCTGTCTAA AGAAATACAG TTACAGGGCT	300
TACAACTTCT TAAACAGACT CTTGAACTT TAGTTGAGCT AGAAAAACAA CGATCTAGTA	360
AGTTAGATTT AATTTCTCGT AAAGAATTAA TGGATCTGCT AGGTATAAGT GCTACAACCC	420
TTGATAACTG GGAGGATCTT GGTCTTAAAC GATATCAGAC TCCGATGGAT GGAGCTAAGA	480
AAGTATTCTA TCGTCCGTCA GATGTGTATT TATTTTTCAG AATAAATAG GAGTTATGAA	540
ATGAAAATTG TTAATTTCAA ACCAACTAAA CAAATAGACG ATGGGTTTTA ACTGCCAGGT	600
ATTGACATTC TATTTGTCTC AG	622

(2) INFORMATION FOR SEQ ID NO: 303:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1929 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

CGCTAACTTG CAAACAAAAG AAGAACGCAA ACTCCACAAA TCCTTTACGC AGAAACTCAA	60
TCTCATCTAC TTACCTTGCT GACTTGGTAG AGTATGTTGC AGACAAAGAC TTCTCAGTAA	120
ACGTAATTC TAAATCAGGT ACAACAACTG AACCAGCGAT TGCTTTCCGT GTCTTTAAAG	180
AACTCTTGGT TAAGAAATAC GGTCAAGAAG AAGCTAACAA ACGTATCTAT GCAACAACTG	240
ACCGCCAAAA GGGTGCTGTT AAGGTTGAAG CAGACGCTAA CGGTTGGGGA ACATTGTGTG	300
TTCCAGATGA TATCGGTGGA CGCTTCTCAG TATTGACAGC CGTTGGTTTG CTTTCAATCG	360
CAGCATCAGG AGCTGACATA AAAGCTCTTA TGGAAGGTGC GAATGCAGCT CGCAAAGACT	420
ACACTTCAGA CAAAATCTCT GAAAACGAAG CTTACCAATA CGCAGCTGTT CGTAACATCC	480
TTTATCGTAA AGGCTATGCA ACTGAGATCT TGGTAACTA TGAGCCATCA CTTCAATACT	540
TCTCAGAATG GTGGAAACAA TTGGCTGGTG AATCAGAAGG AAAAGACCAA AAAGGTATCT	600
ACCCAACTTC AGCCAACTTC TCAACTGACT TGCATCCT TGGTCAATTT ATCCAAGAAG	660
GAATCGTAT CATGTTTGAA ACAGTTGTCC GTGTTGACAA ACCTCGTAAA AACGTGCTTA	720
TTCTACTTT GGAAGAAGAC CTTGACGGAC TTGGTTACCT TCAAGGAAAA GACGTGACT	780
TTGTAAACAA AAAAGCAACT GACGGTGTTT TTCTTGCCCA CACAGATGGT GATGTACCAA	840

1333

ACATGTATGT GACTCTTCCA GAGCAAGACG CTTTCACTCT TGGTTACACT ATCTACTTCT	900
TCCAATTGGC AATTGCCCTT TCAGGTTACT TGAATGCTAT CAACCCATTG GACCAACCAG	960
GTGTTGAAGC TTATAAACGT AACATGTTTG CCCTTCTTGG AAAACCAGGA TTTGAAGAAT	1020
TGAGCAAAGA ACTTAACGCA CGTCTATAAT AGAAGAAAAG AGTGGTTTGC CCACTCTTTT	1080
TACTCTCTTT ATCCATAGAA ATTGGACTCA GCCAAGACTT GTGATATAAT ATAGAAAGCA	1140
AAAAGGCAGA CGCCTAGATA ATAGGAGAAA CTATGTCAAA AGATATCCGC GTACGTTACG	1200
CACCAAGTCC AACAGGACTA CTACACATCG GAAATGCTCG TACAGCATTG TTTAATTACT	1260
TGTATGCGCG CCATCATGGT GGAACATTTC TCATCCGTAT CGAAGATACT GACCGTAAAC	1320
GCCATGTCGA GGATGGTGAA CGTTCACAAC TTGAAAACCT TCGCTGGTTA GGCATGGATT	1380
GGGATGAAAG TCCAGAATCA CATGAGAATT ATCGCCAGTC TGAGCGTTTG GACTTGTATC	1440
AAAAATATAT TGACCAACTA TTAGCTGAAG GAAAAGCCTA TAAATCTTAC GTTACAGAAG	1500
AAGAGTTGGC AGCTGAACGC GAACGCCAAG AAGTAGCTGG CGAAACACCA CGCTACATCA	1560
ATGAATACCT TGGTATGAGT GAAGAAGAAA AAGCAGCTTA CATCGCAGAA CGTGAAGCAG	1620
CAGGGATCAT CCCAACTGTT CGTTTGGCTG TCAATGAGTC AGGTATCTAC AAGTGGCATG	1680
ATATGGTCAA AGGCGATATC GAATTTGAAG GTGGCAATAT CGGTGGTGAC TGGGTTATCC	1740
AAAAGAAAGA CGGTTACCCA ACTTACAAC TTGCCGTTGT TATCGATGAC CACGATATGC	1800
AAATCTCTCA TGTATCCGT GGAGATGACC ATATTGCTAA TACACCAAAA CAGCTTATGG	1860
TCTATGAAGC TCTTGGTTGG GAAGCTCCAG AGTTCGGTCA CATGACCTTG ATTATCCACT	1920
CTGAAACTG	1929

## (2) INFORMATION FOR SEQ ID NO: 304:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 708 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

AAATTTAAGA AAAAGGAGAC ACATCATGTC TAAAAAAGTA TTATTTATCG TCGGATCACT	60
ACGTCAAGGT TCTTTCAACC ACCAAATGGC GTCGAAGCT GAGAAAGCAC TTGCTGGTAA	120
AGCGGAAGTT AGTACCTTG ATTATTCAGC CCTTCTCTC TTCAGCCAAG ATTTGGAAGT	180
TCCAACACAT CCAGCTGTAG CTGCTGCTCG TGAAGCAGTT CTCGTTGCGG ATGCTATCTG	240

1334

GATTTTCTCT CCAGTCTACA ACTTCTCTAT CCCTGGTACA GTGAAAAACT TGCTTGACTG	300
GCTATCTCGT GCCCTTGACT TGTCTGATAC ACGTGGCGTT TCTGCCCTTC AAGACAAGTT	360
TGTCACAGTA TCATCTGTAG CCAATGCAGG GCACGATCAA CTTTTCGCTA TCTACAAAGA	420
CCTCTTGCCA TTTATCCGTA CACAAGGCGT TGGTGATTC ACTGCTGCAC GTGTTAATGA	480
CTCTGCCTGG GCAsACGGAA AATTGGTTCT TGAAGAAACA GTCCTAAACT CACTTGAAAA	540
ACAAGCTCAA GACTTGGTCG AAGCTATCAA GTAAC TAACA CTCAATAAAA ATCAAAAAGC	600
AAACTAKGAA GCTArCCGCA AGCTACTCaa gCACTGCTTT GAGGTTGTAG ATAGAAGTGA	660
CGAGTGThna ACATATATAC GGTAAGGCCA CACTGACGTG GCTTGAAn	708

(2) INFORMATION FOR SEQ ID NO: 305:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 781 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

CTTCTTTTCT TGGAAATAGG TGTATAATAC GTTTATTAAA TTTTGGAGGA GTTGTCTATG	60
AAGAAAAGTT TTATCCATCA ACAAGAAGAA ATTTCCCTTG TCAAAAACAC TTTTACCCAG	120
TATTTGAAAG ATAAGCTAGA AGTTGTCGAA GTTCAAGGTC CTATCTTGAG TAAGGTCGGT	180
GACGGAATGC AGGACAACCT GTCTGGTGTG GAAAATCCAG TATCGGTCAA GGTTCTCCAA	240
ATCCCTGATG CTACTTATGA AGTGGTGCAC TCACTTGCTA AATGGAAACG CCACACCTTG	300
GCTCGTTTTG GCTTTGGTGA AGGAGAGGGT CTCTTTGTCC ACATGAAAGC CTTTCGTCCA	360
GATGAGGATT CCTTGGATGC AACCCACTCT GTTTATGTTG ACCAGTGGGA CTGGGAGAAG	420
GTTATCCCAA ATGGTAAGCG TAACATCGTT TATCTAAAAG AAACAGTTGA GAAGATTTAT	480
AAGGCTATTC GCCTGACTGA GCTAGCTGTT GAAGCCCGCT ATGACATCGA GTCTATCTTG	540
CCAAAACAAA TTACCTTTAT CCATACAGAA GAATTGGTAG AACGCTACCC AGACTTGACA	600
CCGAAAGAAC GTGAAAATGC GATTTGTAAA GAATTGGAG CCGTCTTTTT GATTGGTATC	660
GGTGGCGAGT TGCCAGATGG TAAACCGCAC GATGGACGTG CACCAGACTA TGATGACTGG	720
ACAAGCGAGT CTGAGAATGG CTACAAGGT CTAAATGGTG ATATTCTTGT CTGGAATGAG	780
T	781

(2) INFORMATION FOR SEQ ID NO: 306:

- (i) SEQUENCE CHARACTERISTICS:

1335

- (A) LENGTH: 846 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

```

CCCGCATCTT GTAGGGTTT AACGGGCACG ATTTTCATAT CCGTCTTGAT TGTTTTAGCC      60
GCTTCTAGGG CTGTTTGGA GTTGTTTTTC GCGTCCGGAT GCGCCTTTTG TTCTTCTTCG-----120
CTAACAGGGT TATCAGGAGC AAAGAAAATA GCAGCACCTG CCCTAGCCGA AGCTACAACC      180
TTCTTATCAA TACCTCCAAT GTCTCCACA TTACCATCGC GGTCAATGGT ACCTGTACCG      240
GCAACAATAC GACCATTACG AAGATCTGGG TGAGCTATTT GAGTATAGAT AGCTAGACTA      300
AACATGAGAC CAGCACTTGG ACCGCCAATA CCAGCTGTTG AAAAGCTAAT TGGGACATTG      360
CTGATTACCT CTGTACGGTC AATCAAGCCG ATTCCAATTC CATTTTGGCC ATTTTCCAAG      420
GTGATGATTT TTCCTTCTGC AGACTTGGTT TGCCCATCCT CTTCATAGGT GACCTTGACG      480
GAATCCCCTA ATTTTGGAGA ACTGACGTAA TCAATCAAGT CTTTGGAAC TCAAAGGTC      540
TGATCATGA CTGCTGTGAC TGTATCAGAG ATATTGAGAA TCCCTTTAAA GGTGAATTA      600
TCCGTCACAT TCAAAACATA AACTCCAAAG TACTTGAGTT CGATATCCTT ACCAGCTGTT      660
TTTAGTCCTT GATACTTGGC CATATTTTGC GATGTTTGCA TGTAAGATTG ATTGATTGCG      720
ATAAATTCAA CATCGGAAGA ACCACCTGTA GTCTCCTGAG CACTACGAAT ATCTGTAAAA      780
GGTGTCAACC AAGCATAAAT CATATGAGCT AAAGTGGCAT GTTGAACACC AACCGTAACC      840
AATTGT

```

## (2) INFORMATION FOR SEQ ID NO: 307:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 829 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

```

GCGATCTGCT TGGGCTTTTC CTATTACCTT ATCTAATAAA TAGGTACGCA GACTCATAAC      60
CATATAAAGT CCACCCCCCA TGGCACCGAC AAGAGCTACA TAAAGAAGC TCCACAAACG      120
TCCACTTGGT TGAAGAAAA ATCCTAACAG CCACTGGATG GTTCCTATTA ACAGAAACAT      180
GACTAGGGTC AGCAAACCTGA TTAATGGT TCGCTTCAAA ATCACCTTGC GCTTGACACC      240

```

1336

AGTTACTTTA CAAATATCCC GATACATCAA GACGTTAGGA ATGATGAGAG CAATGGTTGT	300
TGAAATCAAA GGACCATAAC TGTGGAAGAG GGCCATGGTA GGTAGTTGCA AGACTAGCTT	360
GGCAATAGAA CCATAGATAA AATAGAGAAC GGCCTTGCGG TTCCGGAACA TGGCCTGAAG	420
CATTGGAGAC AAGACCATGT ACAAGCCTAA AATAATAGAC TGCAAACTG CAAAGACAAA	480
TAAGCCCAGA GCCAACTAT CTGGCTTACC ATAGAAGACC GTATAAAGAG GTTCTCCTAC	540
CATAACCACT CCAACCGTTG CTGGTAGCAA GAACATAAAG AGTAGGGTGA GACTGTCCTG	600
AACGAGACGA GAAGCTGCTT TCAAGTCCCC CTGACATAG TTTCCGTC AAGTGGCAA	660
ACCAACACTC CCAATCGAAA CCCCTACAGA AATCAAAATC ATCGTGATTT TATTAGGATT	720
GGCTGAGAAA TAAGAAAACA TGACAACCAA GTCCTCATTG CTGTAGTTGG TAAACCAGCT	780
CATACTATTG ATAAAGGTCA GCTGAGTCCA AATCTGGAAG AGCTGGATG	829

## (2) INFORMATION FOR SEQ ID NO: 308:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 464 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

CGAACATCTT GCTGGCTGAT TCGTCTGCCG CCATCGCAGC CCCGAACACA TTGCGACCCA	60
TGGCAAGCGG GCTCAATCCG CACATGGGAT CCGTGCCAAA GCCCCGCGTG TGCATCATTT	120
GCTCATCTAG TAACGTATGA GGTTCGCCTT CGCTGTCGAT AAACCGATAT TCAATCGCAC	180
CACTGCTCGT TCTCCGCGGA GGGGAAACCG ACTGCGGTAG GATGAACTCC AGAGAAGAGA	240
GATCAGGACC TACCAGGTGC GGCTCGTTGA AGCTGTTGCC GCTTAGCAGC AGGCTCGCCA	300
CCACGCATTC CCAGAACTCA ACGGGGGTTT GATCGGCGTT CGGTTGCTGA CTAATAACTC	360
GGTGACGGG ATGCGAAGTG GCCACTTCTG GCACACCGTT CTTGTCTTCG TAGAGAGCAA	420
TTGGGAGGGT GGCCAGCGTT TCGCGATGA GCGCACGCA GGCC	464

## (2) INFORMATION FOR SEQ ID NO: 309:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 982 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:



1337

CCGCTCTATAA TGGTAATAGA TTTTATTTGG AGGTTTTTAT GTCATTTCTA TCAAAAAATG	50
GAGCAGGTAT CTTGGCCTGC CTTCTCATTT CCATCCTATC TTGGTACTTA GGAGGATTCT	120
TCCCTGTGGT TGGCGCGCCC GTTTTGGCCA TTTTCATAGG CATGCTCCTA CATCCCTTTC	180
TCTCGTCCTA TAAACAACTG GATGCTGGTT TGACCTTTAG TTCCAAGAAG TTGCTCCAAT	240
ATGCCGTTGT CTTGCTTGGT TTTGGTCTCA ATATCTCGCA GGTCTTCGCA GTTGCCCAAT	300
CTTCACTCCC TGTCATCCTG TCCACTATCT CAATAGCTCT GATTATTGCC TACCTCTTCC	360
AGCGTTTCTT TGCCCTGGAT AAAAACTGG CTACCTTGGT TGGAGTAGGT TCTTCTATCT	420
GTGGGGGTTG TGCCATTGCA GCGACAGGCC CGTTATTGAT GCTAAGGAAA AGGAAGTAGC	480
CCAAGCCATT TCCGTATCT TTTTCTTCAA TGTCTTGGCT GCGCTCATCT TTCCAACCCCT	540
CGGCACCTGG CTTCTCTAT CCAATGAAGG CTTGCGCCTC TTTGCAGGGA CTGCGGTCAA	600
CGACACTTCC TCTGTAACGG CTGCCGCCAG CGCTTGGGAC AGTCTTTACC AAAGCAATAC	660
CCTCGAGTCT GCAACCATTG TTAACTCAC ACGTACTTTG GCCATTATCC CTATCACGCT	720
CTTCTATCC TACTGGCAAA GTCGCCAACA AGAAAACAAG CAAAGCCTGC AACTGAAAAA	780
AGTCTTCCCA CTTTTATCC TTTACTTTAT CCTTGCCTCT CTCCTCACTA CACTACTCAC	840
CTCTCTAGGT GTGTCCAGTA GTTCTTTAC TCCTCTCAA GAACTCTCTA AATTCCTTAT	900
TGTCATGGAC ATGAGTGCTA TCGGTCTCAA AACCAATCTG GTCGCTATGG TCAAATCCAG	960
TGGAAATCC ATTCATCATG GA	982

(2) INFORMATION FOR SEQ ID NO: 310:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1939 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

CTAGCTGCCA ATATGATTGG GGTGCAGAAG CGCGTGATTA TCTTTAATCT TGGCTTGGTT	60
CCTGTGGTCA TGTTTAACCC AGTGCTTCTG TCCTTTGAAG GATCCTATGA GGCAGAAGAA	120
GGCTGTTTGT CCTTGGTAGG TGTGAGATCA ACTAAGCGTT ATGAAACCAT AAGGCTTGCC	180
TATCGTGACA GCAAGTGGCA GGAACAGACC ATTACCTTGA CAGGCTTCCC AGCTCAGATT	240
TGCCAGCATG AGCTGGATCA CTTGGAAGGA CGAATCATTT AGGAGGAAAG CAAATGAAAC	300
GAATAGTCTT TGAACCTATT TTTATCGCAA CGACCTGGTA TATCTTTTTC CCGCCCTTA	360

1338

ACCTGACCAG CTGGGAATTT CTCTTCTTCC TCTGTGGGCA TTTGTTAGTT GTGGCAATAT	420
TATTTGGCTT TGGCAAGGGG ATAAACCTTG TCAAAACGGT TCATGTGCGC CACGGTAAGG	480
CGGAAGCTGC CTTAAATCTT GAGGGTTTCA AAATCAATCG GTTAGGGAAA ATTCTGTTAG	540
CTTCGATTGG AGGAATTCTT CTCTTGGCAG CTTTGGTTTc CTGGGTAAC TCCAGCATGT	600
TTCAGGCTAA AAATTATGCC AATGTAGTCA CGGTTACGGA AAAAGACTTT ACTGAATTTc	660
CTAAGAGTGA CACCAGTAAG GTTCCTATCC TAGATAGAAG TACTGCTGAA AAAATTGGAG	720
ACCGCTACTT GGGTTCCTTA ACCGATAAGG TGTGCGAATA CGTAGCGGCA GATACCTATA	780
CCCAATTGAC AATTGATGGG AAACCTTATC GGGTCACACC ACTAGAATAT GCAGACCCTA	840
TCAAATGGTT TAACAATCAA GCCAAGGGAA TCGGTGAGTA TATTAAGGTG GACATGGTAA	900
CTGGAATGC GGATTGGTG GACTTGAAGA CACCAATCAA GTATTCAGAC TCGGAGTATT	960
TTAACCGTGA TGTCAAACGT CACCTGCGCT TGAAGTACCC GACCAAAATC TTTAAACTC	1020
CATCTTTTGA GGTGGACGAT GAGGGCAATC CTTTCTATGT AGCAACGGTT TACCAAAAGC	1080
AATTTGGACT TGCTGTTCTT CGTCCTGCTT CAGTCATTAT CTTGGATGCT ACAAATGGAG	1140
AAACCAAGGA ATACAGCTTA TCAGATGTTT CAGAATGGGT GGACAGGATC TATCCAGCAG	1200
AGGAAACCAT TGAGCAAATC AACTACAACG GCAAGTACAA GGACGGTTTC TTGAATGCCA	1260
TGATTTCCAA GAAAAACGTG ACCCAGACTA CCAATGGCTA TAATTACTTG TCTATCGGTA	1320
ATGACATCTA TCTCTACACA GGTGTGACGT CGGCTAATGC GGATGAGAGT AATCTTGGTT	1380
TCATCCTTGA AAATATGCGA ACAGGAGAAA TCACTAAGTA TAGCTTGGCT TCTGCCACAG	1440
AAGAATCAGC CCGTGAATCA GCAGAAGGTG CTGTTCAAGG GAAATCCTAC AAAGCAACCT	1500
TCCCAATCCT CATCAACCTC AATGACAAGC CTCTCTACAT CATGGGCTTG AAGGACAATG	1560
CTGGCTTGGT CAAAGAGTAC GCCCTGGTAG ACGCAGTCCA GTACCAAAAT GTTATCGTTG	1620
CTACTACAGT GGAAGAGATG CTCAGCAAGT ATGCCAATAA AAACGACCTT GAAATTGACA	1680
ATGCAACGAC AGAAAGCATC AATGGAGTAG TAGCAGACCT CAAATCAGCT GTTATCAAGG	1740
GAGACACTGT CTACTTCTTT AAAGTTGATG GCAACATCTA CAAGGTCAAG GCTTCAGTAT	1800
CCGATGACCT TCCTTACCTT GAAAATGGTA AAACCTTCGA AGGTCAAGTA GGAAAAGACA	1860
ATTATCTCAA GACCTTTAAG CTACGGTAAA AATAGGTTTT TTTCAGAAAG TATATGTTAT	1920
AATAAGGTAA ATTAAGCCG	1939

(2) INFORMATION FOR SEQ ID NO: 311:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 907 base pairs
  - (B) TYPE: nucleic acid

1339

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

CCTGCTAATA GAGAGAAAGA CTAGGAGTAG AAGTAAGCCA ATTAAATAAT GAGAAAGTTT	60
CATACCCCGT CCTTTCATGT AGATTGGTA TCGAAAGATA TCTGCGGATA TAAATGTAAC	120
ATTATTTTC TAATCTGTCA ATAAAATTC TGACAATTTA ATAAATACAA CAAGGAGAGA	180
GCAACAAGAC TTTCTCCTTT GTTATCCTAT TCTAAATGT TTTTACCTTA ATCTGATAAA	240
ATAATATCTT CGAGGGAGTA GCTAGCCGTC CAATCAAGAT ATTGTTTAGC TTTTGAAGCA	300
TCTGCTAGGA CACTGGCTGG GTCAGTAGCA CGTCGAGCAA CAATCTCGTG TGGGATTTT	360
TAATTTAGTA ATTCTTCAGC AGTTTAAAG ATTTCTTTGA TAGTATAGCC TTTTITAGTT	420
CCTAAGTTAA AGATTGAGA AGAACTGTCT TCTTGAAATA GGTAGTTCAT TCCTTAACA	480
TGAGCCTATG CAAGGTCCAA GACATAAATG TAATCTCGAA TACATGAACC GTCACGTGTA	540
TCGTAGTCAT CTCCAAATAT TTTAAGCTA TCATTTTGTC CCAATGCGGT CTTGTTGATA	600
TTTGAATGA TGTGAGTTGG ATTTTTCACA CGCAGACCGT TTGAAGCATC CATTTCAGCC	660
CCAGCAACAT TAAAGTAACG GAAAATAACA TATTTCCAGT CGTAGCGATT GGCCATCCAG	720
TAAATCATTC GTTCGCCCAT CAGTTTGTG TCTGCATAAG GGTGACAGG GTCGAGCAGG	780
GTATCTTCAG TCACCGGCTT GTCAATACAG TTATTTCCAT AGAGAGAAGC AGTCGAAGAG	840
AACATGATTT TTTGAATGCC AACTTCAGAT AAGACTTGA GAACTTGTT CATACCAGCA	900
ACGTTGG	907

## (2) INFORMATION FOR SEQ ID NO: 312:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 2170 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

CCACATAAAG GTAAATATCT TTTGTACTAT CTTGGGCATC CAAGAAAAGC AATTGGGCAA	60
TAACAGAGTT AGCCATATTG TCTCAACCG GACCTGTCAG CATAATGATG CGGTCTTTGA	120
GAAGACGTA GTAAATATCG TAAGAACGTT CTCCACGGCT TGTGTTTCA ATAACACAG	180
GAATCATCA TTTCTCCTTT TGAGTTTAA TTTGTTGGT CAAATCACTG AAGATAAGAC	240

1340

TATTATAATA	TCTTGGTCAA	AAAAGGTCAA	ATTTTGGCTC	TGCTTTCATT	AGACAGAAAC	300
AAAAACCCAA	CCTCCTTTCG	TGACTGGAAA	TACTTTTCCA	AGTCATTCTT	CTTTTCGATC	360
TTATTTTGTA	CCGAACAAGC	GGTCTCCAGC	ATCTCCAAGA	CCTGGAACGA	TATAACCGTG	420
TTCGTTCAAA	CGTTCATCCA	AGGCTGCTGT	AAAGATTTCT	ACATCTGGAT	GAGCTTCTTG	480
AAGGGCTTTT	ACACCCCTCTG	GAGCAGATAC	AAGGCAGACA	AATTTGATAT	TTGATGCGCC	540
ACGTTTTTTA	AGAGAATCAA	CAGCCAAGAT	TGCTGAGCCA	CCTGTTGCCA	ACATTGGGTC	600
TACTACAAAA	ATTTGACGTT	GGTCAATGTC	CTCAGGCAAT	TTCACCAAGT	ATTCAACTGG	660
TTGAAGTGTT	TCTTCATCAC	GGTACATACC	GATGTGGCCA	ACTTTAGCAG	CTGGAACCAA	720
GTTCAAGAGA	CCATCAACCA	TCCCGATACC	TGCACGCAAG	ATTGGGACGA	TGGCCAATTT	780
CTTACCTGCC	AATTGTTTTT	GAAGTGTGTT	TGTAATTGGT	GTTTCGATTT	CCACATCTTC	840
TAGTGAAGA	TCACGAAGTA	CTTCATACCC	CATCAACATT	GCAATCTCAT	CTACTAGCTC	900
ACGAAAAGCT	TTTGTAGAAG	TATCTGTACG	ACGCAAGATT	GACAATTTGT	GTTGAATCAG	960
TGGGTGATTA	ATAACTTCAA	TTTTTCCCAT	TTTTGGAATT	CCTTCTTTCA	ATTTATTCTT	1020
CTTATTATAC	CAAAAAACGG	TTTAAAAATC	TTTCTAAACC	ATTTATTTTT	GATAATTTTT	1080
ACATTAGATC	AGCCTCTTTA	AGAGCTGTCT	GTAAGTGTCT	AAGTGGTAAA	TGGGTCAATT	1140
CTGTCCCTTT	TTCTTGATAA	AGGTATTGGG	CGTAGTCGTC	CATTGGGTAC	TGGTTGATAT	1200
AAACCACGCG	CTTGACGCCG	ACCTGAAGCA	ATTGTTTTGT	ACAGTTGAGA	CAAGGAAAAT	1260
GGGTTACATA	GGCTGTAAG	CCTTTGGGAA	CACCACGCTC	AGCACCTTGA	AGGATAGCAT	1320
TGACCTCAGC	GTGAAGGGTG	CGAACGCAGT	GGCCTTCAAT	GACCAAACAT	TCGTGATCAA	1380
TACAATGCTC	AGTCCCTGAC	ACCGAACCAT	TGTAACCAGT	GGAAATAACC	TTATTATCTT	1440
TTACCAGAAT	CGCGCCCACT	TTAGCACGTT	TACAAGTGGG	ACGATTCGCA	ATTAGTAGAG	1500
CTTGGGCTGC	AAAATACTCA	TCCCAGGCCA	GTCTTTTTTC	AGTCATCTCT	TTTCTCCTTT	1560
TTCTCTATTT	TTTAAAAAAT	GGTAAACCTA	AATCTGCAAT	CTTTTCAGCT	GGTACCTTCA	1620
TGCCATCCTT	GATCCATTTT	AGAAGGACAG	AGACGATGGC	TGAGCTCCAG	AAGGAATGAA	1680
GATAAGAGCT	GACACCTTTT	GATTTCCCAT	GGTATTTTTT	TAGAAATTCC	TGCATGGCTT	1740
GGACAAAGAT	TTTTTCCAGA	TGGTAATCCA	AGGCCAATTG	AATTACTCTA	GCTTCCTTTC	1800
TGGCCTCCCG	GAAAAGGTGA	ACCCAAACCA	AATAAAGGTC	TGTCTTTAAA	TCGTAATGAT	1860
GCAGCTGTTT	CATAATATTG	TGGACAGTTC	GTTTAAAGAC	GCTCTCTAAA	ATTTCTCTTT	1920
TGGAGTCATA	ATTGCGATAA	AAGGCCGCAC	GCGAAACACC	TGCACGTTTG	ACCAATTCAG	1980
AAATACTAAT	CTTGGTCAGT	TCCTTTTTTT	CCAAGAGTTG	CAAGAGGGCT	GTTTCAATGG	2040

1341

CTTCTCTGGT TAATAAATTG GATTCTTGGT TTGATTTTCT GAGATTTTCA AGAGACTTTT	2100
CAGAGATTCT ACGTTCAGAC ATAACATTTT CTTTCTACTT GTCACAACAG ACGGATGATG	2160
CTTTTGTTC	2170

(2) INFORMATION FOR SEQ ID NO: 313:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 539 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

ATCTGCACGA ATCAGGGCTT TCTAAGTGAC TATTTCCACC GAAATATTAT TTATATCAGG	60
AGGACATTCA TATGTCACGT TATACAGGAC CATCTTGGAA ACAAGCTCGT CGTCTTGCC	120
TTTCACTTAC AGGTACAGGT AAAGAATTGG CACGTCGTAA CTACGTACCA GGACAACAGC	180
GACCAAACAA CCGTTCTAAA TTGTCAGAAT ACGGTTTGCA ATTGGCTGAA AAACAAAAC	240
TTCTGTTTAC TTACGGTGTA GGTGAAAAAC AATTCGTAA CTTGTTCTGA CAAGCTACAA	300
AAATCAAAGG CGGAATCCTA GGTTCAACT TTATGCTTCT TTTGGAACGT CGTTTGGATA	360
ACGTGTGTTA CCGTCTTGGT CTCGCGACTA CTCGTCGTCA AGCTCGTCAA TTCGTAAACC	420
ACGGTCACAT CCTTGTGAC GGGAAACGCG TTGATATCCC ATCATCCGC GTAATCCAG	480
GTCAAGTGAT CTCAGTTCGT GAAATATCAT TGAAAGTTCC AGCAATCCTT GAAGCAGTA	539

(2) INFORMATION FOR SEQ ID NO: 314:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 667 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

CCGGTTTTGC TCCTTCTCTA CGGCTACGAC GTGATGTATC TCTGATGATA TCCACTGTTT	60
CTGTAGCAGG CGTAGGTGTT TCTGGACCTG CTTGTTCTGC TTTTCTCTT GCCGTCGTAT	120
AGGAAACAGC TACCCTTGTT GGGGTTTCAT TGTATTCTCT TTCAAGTTT TTAGGTCTAA	180
CAGGACCTGG ACCTGGTCTT GATCCACTTT CTTCCGCTGG AGAAGAAGGT ACATCTTGAC	240
TTGGATGACT TGGAACACCA GGAGTTTCTC TTTGAATCTC ATCTGCTGGA GAAGCTGGTA	300

1342

CACCTTGACT TGGGTGAGTA GGCACGGTAG GAGCTTTTCT CATAATCTCC TCTACCGTTG	360
ACAAGGAATC AGCCATGAGT TCTTCAGTTG AAGGTTTCATT TGCAGGAGTG CGAACTACTG	420
CCTCATCTTC TTTCAGAACT TCATCATAGC CTTTTACTTT TTCTAAATCT CTCAGAATCT	480
GCTCTTTAAA GCGTAATTC TCTTCTGCTC TTGACTTTTC ACTCAAAAGT TTTTCCTCCT	540
TGTTGAGAAT CCATAATATT AGAGCTGAGA AGTCCAAAAA AAGCAATCTA TGATACTTTT	600
CCTAACGGAT TTTGTCAATT CCCAGACCAT ATCATACCAT GTTTCCTCTG CAAAGGTTGA	660
CTGGGAA	667

(2) INFORMATION FOR SEQ ID NO: 315:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1483 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

GGGAAGCCAA GGTATTTTAT CGGATGAAGT TGTTACTAGT TCTTCACCGA TGGCTACAAA	60
AGAGTCTTCT AATGCAATTA CTAATGATTT AGATAATTCA CCAACTGTTA ATCAGAATCG	120
TTCTGCTGAA ATGATTGCCT CTAATCAAC CACTAATGGT TTAGATAATT CGTTAAGTGT	180
TAATAGTATC AGCTCTAATG GTACTATTCG TTCCAATTCA CAATTAGACA ACAGAACAGT	240
TGAATCTACA GTAACATCTA CTAATGAAAA TAAGAGTTAT AAGGAAGATG TTATAAGTGA	300
CAGAATTATC AAAAAAGAAT TTGAAGATAC TGCTTTAAGT GTAAAAGATT ATGGTGGCGT	360
AGGTGATGGG ATTCATGATG ATCGACAAGC AATTCAAGAT GCAATAGATG CTGCAGCTCA	420
AGGGCTAGGT GGAGGAAATG TATATTTTCC TGAAGGAACT TATTTAGTAA AAGAAATTGT	480
TTTTTTAAAA AGTCATACAC ACTTAGAATT GAATGAGAAA GCTACAATTC TAAATGGTAT	540
AAATATTAAG AATCACCCCT CCATTGTTTT TATGACAGGT TTATTTACGG ATGATGGTGC	600
GCAAGTAGAA TGGGGCCCAA CAGAAGATAT TAGTTATTCT GGTGGTACGA TTGATATGAA	660
CGGTGCTTTG AATGAAGAAG GAACTAAAGC AAAAAATCTA CCACTTATAA ATTCTTCAGG	720
TGCATTGCT ATTGGGAATT CAAATAACGT AACTATAAAA AATGTAACAT TCAAGGATAG	780
TTATCAAGGG CATGCTATTC AAATTGCAGG TTCGAAAAAT GTATTAGTTG ATAATTCTCG	840
TTTTCTTGGG CAAGCCTTAC CAAAAACGAT GAAGGATGGG CAAATCATAA GTAAGGAGAG	900
CATTGAGATT GAACCATTAA CTAGAAAAGG TTTTCCTTAT GCCTTGAATG ATGATGGGAA	960
AAAACTGAA AATGTGACTA TTCAAAATTC CTATTTTGGC AAAAGTGATA AATCTGGGGA	1020

1143

ATTAGTAACA GCAATTGGCA CACACTATCA AACATTGTCTG ACACAGAACC CCTCTAATAT	1080
TAAAATTCAA AATAATCATT TTGATAACAT GATGTATGCA GGTGTACGTT TTACAGGATT	1140
CACTGATGTA TTAATCAAAG GAAATCGCTT TGATAAGAAA GTTAAAGGAG AGAGTGTACA	1200
TTATCGAGAA AGCGGAGCAG CTTTAGTAAA TGCTTATAGC TATAAAAACA CTAAGACCT	1260
ATTAGATTTA AATAAACAGG TGGTTATCGC CGAAAATATA TTTAATATTG CCGATCCTAA	1320
AACAAAAGCG ATACGAGTTG CAAAAGATAG TGCAGAAATTT TTAGGAAAAG TATCAGATAT	1380
TACTGTAACA AAAAATGTAA TTAATAATAA TTCTAAGGAA ACAGAACAAC CAAATATTGA	1440
ATTATTACGA GTTAGTGATA ATTTAGTAGT CTCAGAGAAT AGT	1483

(2) INFORMATION FOR SEQ ID NO: 316:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2453 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

CCTGAACGCT TTTTATAAA TATCATAAAG CCAATCTGAT TTATCAAGTG TGTCTAAGCG	60
ACGCGAATTA AAATTCATTG CATACTCCAT CGCTTCTAAA AAATCATT TTGAAAAGAC	120
GTAAAAATCA TCTAAATTCT GACTCCAATA TAATAACAA ACCAATCCCA TAATATCCTC	180
TGGTTGATTA TTCAATAAAT TTAAGTTGGT TTCATAAAAC CCTGGAGTTC CAAATAGAGG	240
CAACTTTTTT TCTCAATTT GAGTTTCTTT CCTTAGGGCA TGCTCAAAGT CTATAATATA	300
AATATTATTT CTATTATCAA TAAGTATATT ATTAAATGAT AAATCTCTAT AGGAAAGATT	360
ATATTTGGAG TTTATTATCT CCATATAATC AATTAATGTT AAAAACCAAT CATACGAGCC	420
ACTAACCATA TTATACTCGC TTAATTTATC TGCAATAATA AACTCAAATT CCACAAAATA	480
CGAATTCTTT ATGTAAAAAT CGTTAAAAAC TTTTGGAGTA AATTCCTCCT TTTCCAATTC	540
TACTAATATT TCTCTTTCAT TTATTAAACG ATTCACAGAA TCTCTATTG TAAATCAAC	600
CAACGATAAA TCACTAGCTT CTTTAAATAA AGAATAAACT CGCTTTTGAG TATTAAATAC	660
TTTATAAACT CCACCTTTGG CATTTTATAG AATCACTTCC AAAATAATAT ATTGATCAGG	720
AATAGTGTTA TATCTTGGA TATAGTAATC CCTTATTGGA ACATTACAT TTGAAGGGAT	780
TTTCTTATCT CTTTATCCT TGAAAGTGCT ATCTTTTACG AACTCCCCAT ATCTGTAATA	840
TACAACCTCG CTAAGTTGAA ATCTGAAATC TGATGGTATG TTTACACCCT TTACACCTTT	900

1344

ATACAATATT TCTAATTTGT GTAACAAACG TTGAACTCT TTATTATCTT TTGGATAAAAT	960
TGTAATGAAT TTCCCGACTT GTGAATAACC ATTAAGCCCT GTATTTTGCA AAGAAAGTTC	1020
TTTAATGCTA ACCAAAATTT TGAAATTAT CTCTCTCTCT CTAGAAAATA TAAATCAAA	1080
GAATTTTTTA GCAACCAAAT TAGCATTAA TATTGAAGCG CTCAGGTGTA TTTTAAATCC	1140
CTTAGATTGG GTGATATTAG ACGGCAAATT ATATAACCAA TGTTCATCAC TAAATTTATC	1200
ACTAATTTTA TATTCTAATA ATAAATTATG GTATGCGTCT TCTATTTTCTAG TTTTATAGTC	1260
CAATAGTTT AAATACTTTT CGTAATTCAT ATTAAGAAAT CTCTCCATA AATTTTCTAGA	1320
CCATCATTTA AAGCCAAACA ATTTAAAGCG TGATAATAAA TGTTGATAAT CAATGTAATC	1380
TTCACTCCTC TATTTTGTA TCTCTCACC AATAATTTTA TGCTATATCT ATTTTCTCGA	1440
GGCAATTTAT AGGACTTCAA GATAAAACCA TAAAGAGAT AAGTATTATA ATCTGACAAT	1500
CCAGTTTCAG AATAATTTT TAGAAAAATA TCTAGTGATT CTGATAATTC ATCCGGAATA	1560
ATTCTTTTAA CATCGTATTT ATTTTTCATA TCGGCCCTC TCTCTTAAAA AGCTCACAAAT	1620
AAAATTTTAA ATTTCTATAC AACAATCCGA GAGTAGTCTC ACAATTTGAA CATTTCACAT	1680
CACTCTTAAT ATATAAAAAA TGAATTAATC AGAAACCTCT GACTAAGATT TCCTAATTAA	1740
TTCACTTTCT ATATCATAGT AAGGAATTCT ATTATCCCTA ATTGAAAATT GAAATTTTAT	1800
GTTTTATATA TTAACAATTA TCGGGATTGT AAATCTTGTC TAACAAAATG GCAAGTGCTA	1860
CTATGTGCCC CAGAAGGCGA TGCAACGCTA TTTTGAATTG AAAGAGCATA ATCATCCATA	1920
TCATTTAAGT CACGGATTAG CAATGCTTCC TTCTCTCTC CGACAATTCC AAATTTTCTA	1980
ATTACCTTTT CAGGATTATC AAAAAATTCT CCAACAACCT CCATATTTCC TTGAAGTTCA	2040
TTCAAGAAAG CTTTCATTG ACTACTCATT ATATAGCTCC TTTTCTATTA CTTTATTTGG	2100
AATCAAACT TACTGTGACA TTGGAACAC CTCTATCTA CGCTTTCATA TTGCTGCATG	2160
ACACTTTCAA AATCAAATTG CTAATAATAA TTTTTTAAAG CTTAATTTAG ATTAAATTAC	2220
ATATATCTCA AAAAATTGTT TTGAAATTAG TAAATTAAAA TAGGTTTCTG TACTTATAGG	2280
AACTAGTTAT AAAAATTCTG CCCATCATAA AATATCTATT TAAGTAAAC AAAAATTTTA	2340
TAATTTTTTG ATTTTAAAGT GACTATAATC TCCTATCTAT AAATACCATT CGCAGGACCT	2400
GGATCAATCC CTCTAGCCAT CTTATGAACT TGAGTTCCTC CAGACAGTCC CGG	2453

(2) INFORMATION FOR SEQ ID NO: 317:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1049 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear



1345

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

CCAATTTGAA GGCTCTAAAA CAATGGAAAA GTGCTACACA GATGTGACAG AATTTGCCAT	60
TCCAGCAGTA CTCAAAAAC TTAATTATCA CCAGTTTATAG ATGGCTTTAA CAGCGAAATT	120
ATTGCTTTTA ATCTTTCTTG TTCGCCTAAT TTAGAATAAG TACAAACAAT GTTGGAACAG	180
GCATTCAAAG AGAAGCACTA TGAGAATACG ATTCTCCATA GTGACCAAGG CTGGCAATAC	240
CAACACGATT CTTATCATCG GTTCCTAGAG AGTAAGGGAA TTCAAGCATC CATGTCACGC	300
AAGGGCAACA GCCCAGACAA CGGCATGATG GAATCTTTCT TTGGCATTTC GAAATCGGAG	360
ATGTTTTATG GTTATGAGAA GAACTTTAGA TCTTTAGAAA ACCTTGAACA AGCTATTGTG	420
GACTACATTG ATTATTACAA CAACAAGAGA ATTAAGGTAA AGCTAAAAGG ACTTAGCCCT	480
GTGCAATACA GAACTAAATC CTTGGGATAA ATTAATTGTC TAACTTTTGG GGTGCAGTAC	540
ATTTTTGGTA TATATAAAAT TTGTAGGAGC TATATCTACA ATTTTATATT CCCAGTTTAT	600
GGATGTAAT TACTATATTC ACAATGTTAT CCAGTGTTTT TTCTCTAATA TTAAAGGAGT	660
GTCTGTTC TCGAATAAAT TCTTCAAAGT TTAACCCGTC AACTTGTTCC TGAACAAGAA	720
AATAATCATC CACGATATAA AATTCATCAG TTAAATTAGT AGTATAACTT TTATCGGCTA	780
ATTTTTTTAG CATGTGAGCT TCATTTTTTA TATCATCAAG AGCTGTCCAT TCTCTTCAG	840
CATCATAATT CACAAAAGGT CTTGACTGCT TGATGATTAC TTTTGGCCCG TCCGATTTTC	900
TAATTGCCCG ATAAACATTT CCTTTATTG ATCTCTTAAT AATTTTTTCC ATTTGTATT	960
TATTTATTGC AGAGTCCTTA CTTGAAACTT CACATGTGGT TTGAAAATAA ATCCTTTTTT	1020
CTTCTTCTGA AAATAAATCC ATTTCCGG	1049

## (2) INFORMATION FOR SEQ ID NO: 318:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 776 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

TTAGTTGGTT AGAATCAGAA AATCGCCGAA GTGGTTATTT ATTTTGAAT AAATTTAACG	60
AACCAATTAC AGCAAGAGGA GTTGCTCAAC AGTTAAAAA TTATGCTGAT AAATACAAAA	120
TGAATCCTAA AGTAATTTAC CCTATTCTT TTAGGCATTT ATTTGCTAAG AATTTTTTAG	180

1346

CGAAGTATAA TGATATTGCC TTGCTTGCAG ATTTGATGGG ACACGAAAGT ATAGAAACTA	240
CTCGAATTTA TCTAAGGAAA ACAGCTACTG AACACAAAA TATTGTAGAT AAAATTGTTA	300
ATTGGTAAAA AATAACAGGT GGTCAAACTG ACTACCTGCT ATTTTGTGA TTATGGCTCT	360
TATTATGGGA ATATACCTAT GAATTGGGTT GTTATAAAAA TAAAAGATAT TTTTCAATA	420
AATACAGGTC TTTCTTACAA GAAGGGCGAT TTAAGCATTATAATAAAGG TGTTAGAATT	480
ATACGTGGTG GTAATATTAA GCCTTTAGAA TTTTCTCTGT TGGATAATGA TTAATACATT	540
GATACACAAT TCATCTCCTC TGAGCAAGTT TATTTAAAAC ATAATCAGCT AATAACACCT	600
GTATCAACCT CTTTGAACA TATTGGAAG TTTGCAAGAA TCGAGAAAGA CTATGATGGT	660
GTGTGGCTG GTGGATGTAT TTTCCAATTA ACACCATTCG AAAGTGCAGA GATGATGTCA	720
AAATGTCTAT TATGTAACTT GTCCTCTCCG TTATTTTATA ACAAATTGAA AGCAAT	776

(2) INFORMATION FOR SEQ ID NO: 319:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 658 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

TGCAATGCGG CGGCTGCATA CGCTTGATCC GGCTACCTGC CCATTCGACC ACCAAGCGAA	60
ACATCGCATC GAGCGAGCAC GTACTCGGAT GGAAGCCGGT CTTGTGATC AGGATGATCT	120
GGACGAAGAG CATCAGGGGC TCGCGCCACC GAACTGTTCTG CCAGGCTCAA GGCGCGCATG	180
CCCACGGCG AGGATCTCGT CGTGACCCAT GGCGATGCCT GCTTGCCGAA TATCATGGTG	240
GAAATGGCC GCTTTTCTGG ATTCATCGAC TGTGGCCGGC TGGGTGTGGC GGACCGCTAT	300
CAGGACATAG CGTTGGCTAC CCGTGATATT GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC	360
CGCTTCCTCG TGCTTTACGG TATCGCCGCT CCCGATTGCG AGCGCATCGC CTTCTATCCG	420
CTTCTTGACG AGTTCTTCTG AGCGGGAATC TGGGTTCTGA TGTCGACAGC CCGCCTAATG	480
AGCGGGCTTT TTTTCTCTGA GGCTGGACGA CCTCGCGGAG TTCTACCGGC AGTGCAAATC	540
CGTCGGCATC CAGGAAACCA GCAGCGGCTA TCCGCGCATC CATGCCCCCG AACTGCAGGA	600
GTGGGGAGGC ACGATGGCCG CTTTGGTCCC GGATCAATTC GCGCGACCGG ATCGATCC	658

(2) INFORMATION FOR SEQ ID NO: 320:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1475 base pairs
  - (B) TYPE: nucleic acid

1347

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

CCGGCTTAAT TTTTAGAAAA CGTGGGCAGG GAACCTTTGT TCTCTCTCGT GGCAGCTCAA	60
AAAGAAAATT AATCGTTCCA GAAAGAGATA TCCGGGGACT GACAAAAATA TCTGAAGATC	120
CTCATTCTAC AATTGACTCG AGGATTATTC ACTTCAAATT AGAATTGCA AATGAATTTT	180
TAGCAGAAAA ACTACAGGTC GCTTTGCAGA GTCCAGTTTA TAATATTTAC CGCTGCGTA	240
TTATTGACGG TAAACCTTAT GTTCTGGAAC AAACCTTATAT GAGTACCGAT GTTATTCCAG	300
GTATTACTGA AGATATTTTA CAAAAATCGA TTTACAATTA CATTGAAGGA AAGTTAGGAT	360
TGCATATTGC CAGTGCTACA AAAATCTTAC GAGCTTCTTC TAGTTCAGAA AATGAGCAAC	420
ATTACTTGCA GTCCTTCCA ACGGAACCGG TATTTGAAGT AGAACAAGTG GCTTATTTGG	480
ATAACGGAAC TCCGTTTGAG TACTCGATTA GTCGTCATCG CTATGATTTA TTTGAATTTA	540
ATTCTTTTGC ATTACGACAT TCCTCCTAGG AGAAAATGTG AAAATGAAGC CAATCTTTTA	600
CAGACTCTAG TTTAAGAAAA ATTTAAAACA GGGCAAGAAG GTCCCATCTA TGCTTAAATG	660
GTTTCTCTTT TCTAAATAAG ATGGCTTTAA AAGAGTGATC GTTGATCCA TCATGTTGAA	720
AAATATCTTC GTATAGCTTA TAGAGTAGGT ACTGAAATG TTCACCTGAT CTACTTCTTA	780
TAGTTATTTA GTTTTAAATA GTGTTTCAA CATTCTTACA CTGACGAGAA GTTTTGAAGT	840
CTTTTCTTGT AACACATATA GTATACTGTG GTTAGAATAG TAGACTGTGA CTTCTAACAA	900
ATTGCTAGAA ATGAATTTCA ATCTCCCAAT TTATTTGTTT ATATCTTCTT TTAATATATT	960
AAATAAATTC TAAATCATAA TCATTTAAAA AAATTTTATT TTTTATTTT CATTACGAAT	1020
AATATAGATG AAGGGGAAAG AGTATGAAAA CAGAACTGTT TCTTTTGCTA TTAGTTCAAA	1080
AGGAGAAAAA ATGAAAGTAG AAAATATTTT GTATAGGGTG GATCATCGTA AATTGTTTGA	1140
TAATATTTCT TTTGATACTT CGAGTTCAGA CGTGACATTA ATTACTGGTA AAAATGGTAC	1200
AGGAAAGTCA ACTTTACTAT AGTAGATTGA AACTAGAATA GTACACATCT ACTTCTAAAA	1260
TATTGTTAGA AATCGATTTG ACTATCCTGA TCTATTTGTC CTGTTCTTAT TTCATTTTCA	1320
TATATCTCAA ATTGAGTATG ACGAAGTGGC CTCCCATGTC CTGGGAACGC ACTTTCTTCA	1380
TATTTTTCAT ATTCTTGAAT CCATCGATAA AGACTATTGG GATGAATTTT TAAAGTTGAA	1440
CTAATCATTT TTACAGGATG AGATTTACAG CAGAG	1475

(2) INFORMATION FOR SEQ ID NO: 321:

1348

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 560 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

```

GAAATATATA TACTTCATCT TAATAGTGAG CAAGCTAAAC TTAGCATTTC ATGCCCTCAT      60
ATGGGATGTT CTTTGACTAA ATAATATGAT TATCGAGATA TATCTGGATA AATGAACTAA      120
TAAGTCTGAC GCGTAGACTT ATCAAAGTCA TTGGCATACA CCACTATGAA CTCGTGGGTC      180
TGTTCAAATC CCAACACATT ACCTGAGAAG AAAGTTGCAA TGTTGTTTTT GGTGCGGGTT      240
TGAATTTAAA AAATTTGTTA TGTAAGTACCT AATCTAAGGA ATTAGAACAA TGCCTCTAAT      300
TTTTCTTTAA TACACTGAAA CATTGATGAT TCTGGCTGTA TTTTGAAGAC AGCTCTTCTT      360
TGCTCCTGGA AAATATCTTC AGAAGTTATA TTCTCTATTC CTAACGCTAC TTGAGTTTTT      420
TTTCTAAAAT ATTCTTTTCC GTTGCCATCT TTAGAAAAAT CATAACCTTC CCTATCTACG      480
CTGTTACACA AATTAGCTAA AAAAFACTCT GGGGTTGGGA AAGGAAGATA AGAAaCGTAT      540
TTAGCCCATATACTATAAAG                                                    560
  
```

(2) INFORMATION FOR SEQ ID NO: 322:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 643 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

```

CCGCCCCGCC ACCGCTGCCT ATCCTCGGGA GAGGGTCACC TGGAGTGAAC CTAGAACGAT      60
AGACACGGTG CCGTACGACC TCGTACTACT TTCGCCGACG GCCTCGTCCG TTGTCATCCA      120
CGAACTGATC GGACATGGGT GCGAACACTT CAGAGAAAAA ATCGTTGGAC TCGGTGTCCG      180
GCCTGAGGAA CTACGGGTGG TGGCTTTTCC GAAGAACGGC TCCGGGTTTG ATGACGAGGG      240
TACACCCTCC GAAGAGATTG TACTTGTGGA GAACGGCATT GTGAGGCACG CTGTCAGGGA      300
TCGGGGGACT GGAGGAATGG CGCCTTTTTC CGGTTTGACC AAAGTGGCAT CACATGGTGT      360
CAAACCTGGC TCAAGATGTA CGCATCTCAA GGCGGAAGGG GAATCGTCAC AGGAAGGAGT      420
TACCGGAGTA CCCGCCGAAC GCACCGTTTG GATAGAGCAT TTTTCTGCAG CGAACTACCA      480
TTCAGGTCGA GCCTTTTTCG GGTCTGGCCT TGCCTGGGTA GGCAGCCGAG AAGAACTCTT      540
  
```

1349

ATATCCCTTA ATGCCTTTCA CCATGTCAAT TGATATCTAC GAACTGGCCA GCTTATTGTG 600  
GCATTTAGAC GGTCAAACGG AACGAGCAGC TAGGGTACTG TGC 643

(2) INFORMATION FOR SEQ ID NO: 323:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 780 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

GGTACCCACT CATTCTTGAT GAATTGTGAA CAGTTGCCCT TGGGTCGTTT TCGGAGTTGA 60  
AGTCAAGAAG AGGAAAAAAA CAAAAGGAG AAATACTCAT GGCAGTAATT TCAATGAAAC 120  
AACTTCTTGA GGCTGGTGTA CACTTTGGTC ACCAACTCG TCGCTGGAAT CCTAAGATGG 180  
CTAAGTACAT CTTTACTGAA CGTAACGGAA TCCACGTTAT CGACTTGCAA CAACTGTAA 240  
AATACGCTGA CCAAGCATAC GACTTCATGC GTGATGCAGC AGCTAACGAT GCAGTTGTAT 300  
TGTTTCGTTGG TACTAAGAAA CAAGCAGCTG ATGCAGTTGC TGAAGAAGCA GTACGTTTCA 360  
GTCAATACTT CATCAACCAC CGTTGGTTGG GTGGAAGTCT TACAACTGG GGAACAATCC 420  
AAAAACGTAT CGCTCGTTTG AAAGAAATTA AACGTATGGA AGAAGATGGA ACTTTCGAAG 480  
TTCTTCCTAA GAAAGAAGTT GCACTTCTTA ACAACAACG TCGCGTCTT GAAAAATTCT 540  
TGGGCGGTAT CGAAGATATG CCTCGTATCC CAGATGTGAT GTACGTATG ACCCAGATAA 600  
AGAGCAAATC GCTGTTAAAG AAGCTAAAAA ATTGGGAATC CCAGTTGTAG CGATGGTTGA 660  
CACCAATACT GATCCAGATG ATATCGATGT AATCATCCCA GCTAACGATG ACGCTATCCG 720  
TGCTGTAAA TTGATCACAG CTAAATTGGC TGACGCTATT ATCGAAGGAC GTCAAGGTGT 780

(2) INFORMATION FOR SEQ ID NO: 324:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 624 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 324:

CGGGAAAAAT CAGATTGTGG GTTCAGATAT CGAATTAGCC AAGGCTATCG CAACAAAAT 60  
AGGTGTCGAA TTGGAAGTAT CTCCCATGAG TTTTGATAAT GTACTGGCTA GTGTTCAATC 120

1350

AGGAAAAGCC GACCTTGCCA TATCAGGTGT TTCTAAGACA GATGAACGGA GCAAGGTGTT	180
TGACTTTTCC ATTCCCTACT ATACTGCAAA AAATAAACTC ATTGTCAAAA AATCTGACTT	240
GACTACTTAT CAGTCTGTAA ACGACTTGGC GCAGAAAAAG GTTGGAGCGC AGAAAGGTTT	300
GATTCAAGAG ACGATGGCGA AAGATTTGCT ACAAATTCT TCCCTCGTAT CTCTGCCTAA	360
AAATGGGAAT TTAATCACAG ATTTAAAATC AGGACAAGTG GATGCCGTTA TCTTTGAAGA	420
ACCTGTTTCC AAGGGATTG TGGAAAATAA TCCTGATTTA GCAATCGCAG ACCTCAATTT	480
TGAAAAAGAG CAAGATGATT CCTACGCGGT AGCCATgAAA AAAGATAGCA AGAAATTGAA	540
AGAGGCAGTT CGATAAAACC ATTCAAAAGT TGAAGGAGTC TGGGGAATTA GACAAACTCA	600
TTGAGGAAGC CTTATAAGCA TCCA	624

(2) INFORMATION FOR SEQ ID NO: 325:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1237 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

TCTTATGAAG CCGAAGCGTG ATTTATGGCG GATAGGTTTG GTCTGCAGAA AGTGACAAAT	60
CTAGTGCCAT CAGCGTATAT GGAATCTnTG GCTGAGAAAC AGTCCCGGGG TGAAGTGAAT	120
TATGAGCAGG TTTATGAGGA TGCAACGGCT TATCATCATA CCATTGATGC GAGTACAGAG	180
GAGGCAGACT TGGTTTCTCT ACGTATTGTA GAACTATTGT CTCGAAGAGG CTTTAGCTTC	240
AGTCCTGCGA TCTTACTTGC TATTCATAAG GAGTTGTTTC AAGATATATT TGAACCTCG	300
ATTCCGGTAG GTCAATTTTCG TCAGACTAAT ATCACAAGA ATGAACCTGT TTTGAATGGT	360
GAAAGTGTG TGTACTCTGA TTAATCCATG ATTCAAATGA CCTTGGATTA TGATTTTAAT	420
CAGGAAAAAC AAGTTGCATA TGGGACACTA ACCCAGGCGG ATATGGTTAA AAAATCCAG	480
CATTTTATTT CAGGAATCTG GCAGATTCAT CCATTTCGCG AAGGAAACAC TCGGACGGTA	540
ACGGTATTTT TGATTCAGTA TCTTCGTGAG TTTGGTTTTG ATATTGATAA TACACCATT	600
CAGCAACATT CCAAGTATTT TCGTGATGCC TTAGTGTTAG ATAATGCAAA GATTTTACAG	660
CGACGTCCTG AGTTTTTAAC AGCTTTTTTT GAAAATCTCT TGCTCGGTGG TCAAAATGAT	720
TTGTCTTCAG AAAAAATGTA TCTAGATTTA GACCTCGATC TTTTATAATC CTAATACTGA	780
GTAAACATTG AATTTTAGGA AAAAAATGAAG TAAATATTCT CACAAGAAAA CGTATATCAT	840
CAAAGTTTGG CTCTTTGTCA ATTGTAGTGG GTTGAAGAAA AGCTAAGTTC GAGAAAGGGC	900

1351

AAATTTCCGGC CTTTCCTTTT TGATGTTTCAG AGCGATAAAA ATCCGGTTTT TTGAAGTTTT	960
CAAAGTTTCG AAAACCAAAG GCATTGCGCT TGATAAGTTT GATGAGATTA TTGGGCGCTT	1020
CCAGTTTGGC ATTAGAATAG TGTAGTTGAA GGGCGTTGAT AACCTTTTCT TTATCTTTGA	1080
GGAAGGGTTT AAAGACAGTC TGA AAAAATAG GATGAACCTG CTTAAGATTG TCCTCGATAA	1140
GTTCGAAAAA TTTCTCCGGG TCCTTATTCT GAAAGTGAAA CAGCAAGAGT TTGAAGAGCC	1200
GATAGTGATG TATCAAGTCT TGTGAATAGC TCAAAAG	1237

(2) INFORMATION FOR SEQ ID NO: 326:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 461 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

TTTGATTTTT CTGAATTAGA AGAGATTGAA TTGCCTGCAT CTCTAGAATA TATTGGAACA	60
AGTGCATTTT CTTT TAGTCA AAAATTGAAA AAGCTAACCT TTTCTCAAG TTCAAAATTA	120
GAATTAATAT CACATGAGGC TTTTGCTAAT TTATCAAATT TAGAGAACT AACATTACCA	180
AAATCGGTTA AACATTAGG AAGTAATCTA TTTAGACTCA CTACTAGCTT AAAACATGTT	240
GATGTTGAAG AAGGAAATGA ATCGTTTGCC TCAGTTGATG GTGTTTTGTT TTCAAAAGAT	300
AAAACCCAAT TAATTTATTA TCCAAGTCAA AAAATGACG AAAGTTATAA AACGCCTAAG	360
GAGACAAAAG AACTTGCAATC ATATTCGTTT AATAAAAAAT CTTACTTGAA AAAACTCGAA	420
TTGAATGAAG GTTTAGAAAA AATCGGTACT TTTGCATTTG C	461

(2) INFORMATION FOR SEQ ID NO: 327:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1436 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

TAACATTTAG GTACCTCTTC TTAACAAAGT TCAATAGTAA CAATTAATAT TTAAACAAT	60
ATATCAAACA TCAATGACTA GAATACTTGC ATCATCCTTC TTTCCATAGA TTGGATCAAT	120
AGCAGAAGAA TTAAATCTCA TCTTAATTAA CTCTTCAAAA GTTTTATTTT GATTATTTTG	180

1352

ATAGAATTCA TAAAAGCCAT CGCTCATTAA AACAAATTTGT TCACTAGTAA CATCTATTTG	240
ATTAATAATA GCATGGTCTA AAAATCTCTC ATCCAACGAA CCTATCCAGT ACCCACTCGG	300
TTGATTAGAT AATTTTCTGA TTTTTGTAA AATAATTTT TTATTTAAAA CACTATTTGT	360
ACCAATTGAA TCTTTTATCT CATTTTCCC TTTTCAAAT AAGTTATCTA CTCTATGATC	420
AGTTATTTCC ATTTCTTTA CTAACATGAC GCAGTCACCT AGCATCATAT ACTCCAACCT	480
TTTTTCTGAA AGTTTAGCAA ATATTGGTAA GCGATAATAT AGTATATTGA AACTAGAATA	540
GTACACCTCT ACTTCTAAAA CATTTGTTAGA AATCGATTG ACTGTCCTGA TTGATTTGTC	600
CTATTATTAT TTCATTTTAC TATACTCTGT TAATTTATAT GAGTTTAAAC CGATTTCATC	660
TTTAACCTCG AGTAAAGCAG TTTCAAATAT TTGTTAAGA GTTTTGTATT CTTTACAATT	720
AACCGACAAA CTTTCTGATA AAATATGTAC AACTTCTGAG ACTGAATAAC CTATCTCCTC	780
TTTAGAATTA TATAAATCTG TAGCTCCACC AATAATCCAA AAATACTGAT TTTGTGAACC	840
TACAATATCC TCATTTTCTA CGGAACCTCC TTGTATCGAA CAAATTTTAT TTATCTTTAC	900
CATAATACTT CAACCTTTT AGTGTCAAAA GTAAACCAAT TCCTGTCACT GTTAAGAATA	960
GTTCATAAT CTTATTCGAA CCAGTCTTTG GTAATTTTGT TTKACATCT ACTATCTCTT	1020
TAGATTTATT AATATGATTT TCAGTTTCTC TGCCATCTCC AACTATTTTA TAGTTTACTT	1080
CTTCTGTCTT ATTATCTTGT TTATTGTCGA TCTGTCACTT CATTTGTCTA TTATCTTTAC	1140
TTGAGTTAAA CTCTCCGTTT TTTCTGTTAC TATCAATTAC ATTATTTGAA TTAGATTGTT	1200
TTTCTCTTT GTTTTTTCT TTTTCTTTT TATCACTTAA ATTATTTGTT ACAATTTTGT	1260
AAAGCCCATT CTCCGTTACA ATATTGAAAT TACCATCGCT ATCACGTATA ACAGGTTCTT	1320
TCCCATTTGC ATTAGATTTG ATGAATGATA TATACTTACC GGATAAATTA TAAAATTGGT	1380
TATTTAAAAC GGTATTTTA CCCTTTGAAT CCTCAATAAC AATTCCTTCT TTACCC	1436

(2) INFORMATION FOR SEQ ID NO: 328:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 646 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

CCGGCAGACA GGAGAAGGTG TAAATATCA ATCTCAAATG GTTCGTCAAT GGTTCCTGAT	60
ACGTATTTTC CGTCTTTCTT CCGTTGCTTG ACACACTCTG TGAGGAGATA TTCGATTTGC	120
CCATTGACTG AACGAAAGTC GTCTTCTGCC CATGATGCCA GTGCAGCGTA TAACTTTGTT	180



1353

GAGAGTCGAA GGGGGATCTG CTTTTTTTGA GCTTCAGCCA TCTTTAGTAA AGGCTTCCTG 240  
 TGTGACAAT TGGTTGTGCA TCATGATTGC CACAAAGAAC GACAAGGAGA TTTGAAACCA 300  
 TGGCAGCTTT TCGTTCCTCG TCAAGTTCTA CCAATTCCCC TTCATTGAGC CGTCTAGTG 360  
 CCATTTCAAC CATTCTTACA GCACCATCTA CAATCATCTT CCGTGCATCA ATAATGGCAG 420  
 ATGCTTGTG GCGTTGAAGC ATAACGGCAG CAATTTCTGG AGCATAAGCT AGGTAAGTGA 480  
TACGTGCTTC AAGGATTTCC AAGCCAGCAT CCTCAACACG ACTTTGGATT TCTTCACGAA 540  
 TACGGGTAGC AACAATTTCC CTAGAGCCAC GGAGACTACC TTCATCTGCG TGCCCATCAC 600  
 CCGGAGTATC CACATTAGGA GACACATCGT AAGGATAGAT GCGGAC 646

(2) INFORMATION FOR SEQ ID NO: 329:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1653 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

GTTGCAGGTG CAGTAGGTGT TACTTCAGAT ACATTTGAAC GTGCAGAGGC TCTTTTGTAG 60  
 GCAGGAGCGG ATGCGATTGT TATTGATACT GCACATGGTC ATTCTGCAGG TGTCTGCGT 120  
 AAAATTGCCG AGATTGCTGC TCATTTCCCA GATCGGACTT TGATTGCTGG AAATATTGCT 180  
 ACTGCTGAAG GTGCACGTGC CCTTTATGAA GCGGGTGTAG ACGTTGTAA GGTGGTATT 240  
 GGACCAGGTT CTATCTGTAC TACTCGTGTG ATTGCTGGTG TTGGTGTTC GCAAGTAACA 300  
 GCTATCTACG ATGCTGCAGC TGTGCGCGC GAATATGGTA AAACGATTAT TGCTGACGGT 360  
 GGGATCAAGT ATTCTGGAGA TATTGTAAAA GCACTTGCTG CAGGTGGAAA TGCTGTTATG 420  
 CTTGGATCTA TGTGCTGG AACTGATGAA GCTCCAGGCG AAACGAAAT CTCCAAGGA 480  
 CGTAAATTCA AGACTTACCG TGGTATGGGA TCAATTGCTG CTATGAAGAA AGGTCAAGC 540  
 GACCGTTATT TCCAAGGTTG TGTCAATGAA GCAAACAAGC TTGTTCCAGA AGGAATTGAA 600  
 GGTCGTGTTG CTTATAAAGG AGCGGCAGCT GATATTGTTT TCCAAATGAT TGGTGGTATT 660  
 CGCTCTGGTA TGGGTTACTG TGGTGCAGCT AACCTTAAAG AACTACACGA TAATGCTCAA 720  
 TTTATTGAAA TGTCTGGTGC TGGTTTGAAG GAAAGCCATC CTCATGATGT GCAAATTACT 780  
 AATGAGGCAC CAAATTATTC TATGTAAAAA ACAATGAAAA GAACTCCAGT GAAACAGGA 840  
 GTTCTTTTAC AATGTTGTCA ATTTCCATTT ACAGCAGCTT TACCATCCTG AATAGTGAAG 900

1354

ATACTTAGAT TTTCTGGCAG ATTTTGAAGA TGGTCTAAGC TTGTTGTTGT GATAAAGGTT	960
TGGATTGATT GAGAAATCGT TTCTAATAAT TTAACTGTC TAGTGTGTC AAGTTCACCT	1020
ATCACATCGT CAAGCAGTAA TATAGGAGAT TCTGTGGTAA TGCTTTCCAT TAATTCGATT	1080
TCTGCTAATT TTATCGAGAG GACGAGACTA CGATGTTGAC CTTGGCTTCC GAACTAGCA	1140
TCCATCCCAT TTATATAAAA AGAAATGTCA TCTCGATGAG GACCGACACC AGTATTCTTT	1200
TTAAATAAAT CTCTGGATCT ACTTTTTTCT AAAGCAATTT TGAAAGATTC GGATAAGTTT	1260
TGTTTGTGAG TTATATTGAC AGAAGATTGA TAGGATATTG ACAACTCTTC GATCTGATTA	1320
GAGAGTTCAA AATGTTTCTT ACGCCCAAAT GATTCTAGTT TTTTATGAA ATCTAAGCGG	1380
TGATTCATTA CACGACATCC ATAATCAACT AGCTGATCAT CTAACACAGA AAGGAATGTT	1440
TCATCTATTT TTTGAGCTGA TTTTAGGTAA GTGTTTCTTT GCTTTAGGAT GTGGTTATAA	1500
TTGGTTAAGT CAGATAAATA GATTGGCTTA ATTTGCCCAA GTTCCATATC AATGAATTTT	1560
CGTCGAATCG AAGGTGCTCC TTTAATTAGT TGTAATCTT CAGGAGCAAA TAAGACAACA	1620
TTCATGTGTC CTACATAATC TGAAAGGCGT GCC	1653

## (2) INFORMATION FOR SEQ ID NO: 330:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1340 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

GAAACACTGT ATTTCAAAGC ATTTTGTGTT AGTTTAAAAT TACTCCCATT CTCTTTTCC	60
AAACGTACAA TATATCCAAA ACCATTCAAA ATACTAGATT CTATTTTTTA TAATATCACT	120
AAATCCACCT AATTATAGGA CGTTTTGAGA TTTTGTAGTCC CAGTCCCAGT ACCGGAGAAA	180
TATTGTTTTA ATATAATATC TCTTTTGTG TTCTAAGCTC TTAAAAGCAA AAGAACAAGT	240
AAAGAGTCAA GACAAGGATA AAAAGTCCAT ATTAGGGCAA ATAAAAGCT TTAAGACAGA	300
TGACAAATCT AAGTCAAATA AGAAAGACCA TAGCAAAGGT GCAGAGAGAT AAATATTGGC	360
GGTCTTCGGA CTGCCTTTAT TTTTATATCC ATTTTCAAA TCAAATTTAT TCAGACTATA	420
TATGCACATA TACACTTAAA TTCATATAAA AACATGGCTT GTAAAAAATT ACTTTAATCA	480
CAATAATCGC ATTTAAAATT GTGATGTTTG CAAGCTAAAT TACGGACTTC ACTTGAAGT	540
TTCCCTTGT ATCTTTTATA ATAGATAGAA AATTGCTGG CAGATGAATA TCCAACAGAT	600
TCTGCTATCT CTTTATAGG TAGTTCAGTG TTTAAAAGAA GAGTTTCAGC TACATTCATT	660

1355

CTTTTTCTTT GAGTGTACTC TGTAATGCTT TGACAATATT TTTCCTTAAA TAAATTTTTT 720  
 AATTTAGTAC CACTCATTTT AGATATTTTT TCAAGCGTGC CTGATTAC ATTTCGTTGCA 780  
 AAATGATCAT CTAAGAATCT TGCTACATCT TCAAGTCTT TATCATCATC AATTCAATT 840  
 TTATATTTTT TTCTATTTAA GTATGTGTCA ATTACTATAC TTATCCATTC ATTTCGCTTT 900  
 GCTTTAAAGA AAAAATCAGC GGCAGGAGCG TCCATCTTAC AATTTAATAT TTCCATTGCC 960  
 ACTCTTTCTA AGGCCTTTGT AAGTATTATT TGATTCCGTT GAAGCAAGGT TGAATAAAAA 1020  
 GATTCTGGAT TAATGTTAAT AGATGCTAAA TGTMTTCTA TTAGCTCTTT TTTAAAACCM 1080  
 ATGGAACAG CAAGATAACA ACAATTCTCG TGTAATAAAA AAACAAAATT ATCTTTTATA 1140  
 TTATCAAAAT CAAAAGTACA TAGAGAGTTT GCGGTAATAG TTTGATACGG ATTAAACTTT 1200  
 TCTCCGTTTG CACTGACAAT GAACTTGAA TAAATTGAAA CATAGTCTGA CATACTATAA 1260  
 GTGCTATTTT GAACTACTTC CTCTTTGATA TAAAAATCAT GTATATCGAT AATGAAGATG 1320  
 CCTCCTTCAT AAAACCGGTA 1340

(2) INFORMATION FOR SEQ ID NO: 331:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 607 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

TATGTTCTGT ATGAGTTTTT AAGTAGGAAA AACGTGCTAA CCTCTCAGAT TTGGAACTT 60  
 GTAAAAGAAA CTCTTTTTC ACCCGTAGTA GTTGATAATG GGTTCGATCC GGCCTTATTT 120  
 GAAATTGAGA AAAACAATT GCTAGCAAGT TTAGCAGCTG ATATGGATGA TTCTTTTAT 180  
 TTGACATA AAGAATTGGA TAAATTGTTT TTTCATGATG AACGTCTTCA ATTGGAATAT 240  
 AGTGATTTAC GAAATCGTAT TTAGCTGAA ACTCCACAAA GTTCTTATTC TTGTTCCAA 300  
 GAATTTTATG CCAATGATCG AATAGATTC TTTTCCTAG GTGATTTTAA TGAGGTTGAA 360  
 ATTCAAAATG TATTAGAATC ATTTGGCTTT AAAGGTCGAA AAGGAGATGT GAAGGTTTCC 420  
 TATTGTCAAC CTATTCTAA TATCCTTCAG GAAGGTATGG TTCGAAAAA TGTGGGACAA 480  
 TCCATTTTGG AATTAGGTTA TCATTACTGT TCTAAATATG GTGATGAGCA ACATTTACCC 540  
 ATGGATTGAA TGAATGTTT ACTTGGTGA TTTGCTCACT CTAAGCTCTT TACAAATGTC 600  
 CGGGAAA 607

1356

## (2) INFORMATION FOR SEQ ID NO: 332:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 900 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

```

TAAAAATACC GAATTTTGTT TTGTCCTCTA TTTCAACATT GTGAATCGCC TCAGGCAGAG      60
AACCGATACT AAAGATATAA CCAAATAGT  TGTCATTTGC TTTACCGATA TCAATCTTAT      120
TGGTTAAATC AAAATCCAGT TCGTCAATTG CGCCATCGAT GTCTTGATTG ATTTCCAAAA      180
GTTTGTGAAT GAGGTTACCC GTACCGCCTG GGATAATCCC TAACCTAGGA ATGTAGTCTC      240
TCTCATCAAT ACCTGAAATG ACTTCATTGA CAGTTCCATC TCCACCAAAC ACAACCACTG      300
CATCATACTG CTCACGAGAA GCTTCTTCAG CAAATGTGT TGCATCCAGC GCTTTTTCGG      360
TAATTTTGGT TTCAACATAT TCAAAGTATT CTTTTCCTTT ATTCTCCAGC TTTTCTTTGT      420
AATCCAAAGC CTTCTCGCCA CCAGAAGTAG GTTTGATAAT TACCATTGCT TTTTTCATTG      480
ATTTTATCCT TAATTTTAAA CAGAAATGTT TACATTTCTG CGTATGCAAG TAAATGTAAT      540
CCTATTATAC AATGAAAATA CAGAAAAGAG AAATCTGACG TACTGGAGAT TAATACGCTT      600
TTATTCTATT TTCCCATCGC CTAACATCAT CCTTTAAGGG TTCATCCAAG TAAGAATAGG      660
CCTTATCCTT GATCCAATCA GGAATACCGT AAGCTGCCTC TGCTAWGCTA CAAGTGATTG      720
CTGCGAGAGT ATCACTGTCT CCACCAAGTG AGATGGCATT TCTTATCGCA TCTTCGAAGT      780
CTCTACTTTC AAGAAAGGCG ATAATGGCTT GAGGGACAGT TTCCTGACAT GTTTCGTTAA      840
AACGATAGTT AGGACGGATT TCATCTAAAG TTTGAGATAG ATTGTAATCG TATTCTTTTT      900

```

## (2) INFORMATION FOR SEQ ID NO: 333:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 533 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

```

CCTTTCTGGC ACACTGGTCT TGAATACGG CAAAACCTCT GAAAATATCT ATGCTGGAAT      60
GGACGAGGAA TACCGTCGTT ATCAGCCTGC CATCATCACT TGGTACGAAA CAGCCAAACA      120
TGCTTTTGAT CGCGGACAGA TTGGCAAAT ATGGGTGGAA TCGAAAACGA CCTCAAGGGC      180

```

1357

GGTCTCTACA GCTTTAAATC CAAGTTCAAT CCGACCATTG AGGAATTCCG TGGTGAGTTC 240  
AACCTGCCAA CTAATCTCTT TTACCACCTC TCCAATCTGG CCTACACTCT CAGAAAGAAA 300  
CTGCGCAGaa GcATTAAACAG AAAGGAAGCC TATGACCTTT AAACCTTCTCA GCCAAGAAGA 360  
ATTCATCCAG CATACCTCAG CTAGATCCCA ACGCTCTTTT ATGCAGACCG TAGAAATGGC 420  
AGAGCTGCTG AGCAAGCGTG GCTTCAGTAC CCAGTATGTC GGCTACACTG ACCCACAAGG 480  
GAAGGTAGTG GTGTCAGCTG TCCTCTACAG CATGCCTATG ACTGGTG GCC TTC 533

(2) INFORMATION FOR SEQ ID NO: 334:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 544 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 334:

CCAGCAAACCT AGGAAGCTAG CCGTAGTTGC TCAAAGCACA GCTTTGAGGT TGTAGATAAG 60  
ACTGACGAAG TCATGTACAA AACACTGTTT TGAGGTTGCA GATAGAACTG ACGAAGTCAC 120  
TCAAAACACT GTTTTGAGGT TGCAGATAGA ACTGACGAAG TCACTCAAAA CACTGTTTTG 180  
AGGTTGCAGA TAGAACTGAC GAAGTCanna ACCACACCTA CGGCAAAGTG AATCTGAAGT 240  
GGTTTGAAGA GAGTACAACT TGTCTTTTAG AAAAGGAGCC TATAATGAAA GTCTTTCAGC 300  
ATGTAAATAT CGTGACTTGT GATCAAGATT TCCATGTTTA TCTTGATGGA ATCTTAGCAG 360  
TCAAGGATTC TCAAATCGTC TATGTCGGTC AAGATAAGCC AGCGTTTTTA GAGCAAGCTG 420  
AGCAGATTAT AGACTATCAG GGAGCTTGA TTATGCCTGG TTTGGTCAAT TGTCACACCC 480  
ATTCTGCAAT GACAGGTCTG AGAGGGATCC GAGATGACAG CAATCTCCAT GAATGGCTCA 540  
ATGA 544

(2) INFORMATION FOR SEQ ID NO: 335:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 349 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 335:

CCAGGAACTC AAATGTAAGT AGGGGTCCT TTTTGTATA TTTTCAAAT AACGCCTCTA 60

1158

CACTATTTGT AGCAAATTCA CCAACTACAG TTGTATCTTA GTTAAAATAA GTTAGAATAT	120
GTAAGTGAGT ACCAGATATA CCAAGACATC GTCACCATCT AAGGTATATT CAAAATACAA	180
AAGTTGACCA ACTAGATTTT TGAATATCCT TATATATCCA TTCTTAAAT TGGTTTAAAT	240
AGCGTAGTCT TTAAACTAG TTTTGAGAAT CCAAAAAATC TTCCTACATA TGTAAGAAGA	300
TTTTTTAGTT CAGAATGATT AGATTAGCT AATGGATACC TATCTACC	349

(2) INFORMATION FOR SEQ ID NO: 336:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1206 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 336:

CTCCGATAAC CACACCAGCA ATGGAAATAA TTCCATCGTT AGCATCAAGA ACACCCGCAC	60
GCAGGATATT TAAACGACCT GCAAAATTG AATCAATTTC GTGATTGTT TCTGACGCTA	120
AATTTCAAGT TCAAGTTAGC CATCAAGAAG TCTTCTCTGG GTGACTTGTA GTCCAAGCAT	180
TTTTTAGGAT AGTTGTTAAT CCACTTTTCG ATGAATGCGA CTTCTTTGGG AGTCATTTTC	240
TTGGTTCCCT TAGGTAACCA TCTACGAATG AGCCTGTTGT GATTCTCATT AGTTCCCGGG	300
ATCCTCTAGA GTCGACCTGC AGGCATGCAA GCTTGCCACT GGCCGTCGTT TTACAACGTC	360
GATGACTGGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC CTTGCAGCAC ATCCCCCTTT	420
CGCCAGCTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC AGTTGCCGAG	480
CCTGAATGGC GAATGGGGCC TGATGCGGTA TTTTCTCCTT ACGCATCTGT GCGGTATTTT	540
ACACCGCATA TGGTGCACTC TCAGTACAAT CTGCTCTGAT GCCGCATAGT TAAGCCAGCC	600
CCGACACCCG CCAACACCCG CTGACGCGCC CTGACGGGCT TGTCTGCTCC CGGCATCCGC	660
TTACAGACAA GCTGTGACCG TCTCCGGGAG CTGCATGTGT CAGAAATTTT CACCGTCAATC	720
ACCGAAACGC GCGAAACGAA AGGGCCTCGT GATACGCCTA TTTTATAGG TTAATGTCAT	780
GATAAGGATG GTTCTTAGA CGTCAAGTGG CACTTATCGG GGAAATGTGC GCCGAGACCC	840
TATTTGTTTA TTTGTCTAAA TACATTCAAA TATGTATCCG CTCGTGAGAA AATAAACCTG	900
ATAAATGCGT CAATAATATT GAAAAATGAA GAGTATGAGT ATTCTACATT TCCGTGTCCG	960
CCTTATACCC TTTTGTGCGG CATGTTGCCT TCCTGTTTTT GCTCACCAG AAAACGCTGG	1020
TGAAAGTTTA AGATGCTGAA AAATCATTTG GGTGCACAAC TGGGGTTACA TCCAAGTGA	1080
ATCTCCATCA GCAGTTAAGA TCCTCTGACA GTTGATACAG CCGCAAGAAC TATTCCTGAT	1140

1359

GAATGAGCAA CTTTTAAAG TCCTGCGAAT GTTGGGGCGG TAATAATCCC CGTGTTGTAG 1200  
 GCCCGG 1206

## (2) INFORMATION FOR SEQ ID NO: 337:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 813 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 337:

CTGCTCAACT CAGACAGTCA AATTCTGAC TTTACCAAAA GAACCATCAA AAAAGTTGCT 60  
 GAAAAAGGCC ATCAGGTTAT TATTACGACA GGTGCGCCTT ACCGTATGTC AAAAGATTTT 120  
 TACCGTGAAC TGGGCTTAGA CACTCCTATG ATTAACCTCA ACGGATCCCT TACTCATTTA 180  
 CCAGACCAAG TTTGGGATT TGAAGAGTG TTGACTGTAG ACAAAAAATA TCTGCTAGAT 240  
 ATGGTTCAAC GTTCAGAGGA CATTCAAGCC GATTTTATCG CTGGAGAATA TCGTAAAAAA 300  
 TTCTACATTA CAAATCCCAA TGAAGAAATT GCCAATCCCA AACTATTTGG TGTAGAAGCT 360  
 TTCCAGCCTG AAGATCAATT CCAGCCTGAA TTGGTGACCA AGGACCCTAA CTGTATCCTC 420  
 TTGCAGACTA GAGCCAGTGA CAAATATTCC TTGGCAAAAG AAATGAACGC CTTCTACCAG 480  
 CATCAACTTT CTATCAATAC CTGGGGAGGT CCGCTCAATA TCCTTGAATG TACCCCAAAA 540  
 GGTGTCAACA AGGCCTTTGC TTGGACTAC TTGCTCAAGA TAATGAATCG TGACAAAAAA 600  
 GATTTGATTG CCTTTGGAGA TGAACACAAT GATACCGAAA TGCTCGCTTT TGCTGGGAAG 660  
 GGTTATGCCA TGA AAAATGC CAATCCAGAG CTA CTCCCTT ATGCAGATGA GCAAATTTCC 720  
 CTTACCAACG ACCAAGATGG GGTGCGCAAA ACCCTACAAG ACTTATTCTT ATAACCTATA 780  
 CTGATACTCA ATGAGGGGCA AAGAGCGAAC TTA 813

## (2) INFORMATION FOR SEQ ID NO: 338:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 683 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 338:

CCTAGATAAA TGATATAATT CTATTATTGT TCGTAAAAAT TAAAAGGAGA TTGATGATGG 60

1360

ACAAATTATT TAAACTAAAA GAGAACGGTA CAGACGTTTCG TACAGAGGTT CTCGCTGGTT	120
TAACAACTTT CTTTGCAATG AGCTATATTC TCTTTGTAAA CCCACAAATA CTTTCACAAA	180
CAGGAATGCC TGCTCAGGGC GTCTTCCTAG CGACGATTAT TGGTGCAGTA GCGGGTACCT	240
TGATGATGGC TTTTATGCT AACTTACCTT ATGCCCAAGC GCCAGGTATG GGAATCAATG	300
CCTTCTTTAC CTTTACAGTT GTATTCGGGC TTGGTTATTC TTGGCAAGAA GCCCTAGCTA	360
TGGTCTTCAT CTGTGGGATT ATTTCAATGA TTATTACCTT GACAAATGTT CGTAAAAATGA	420
TCATTGAATC GATTCCCAAT GCTCTTCGCT CAGCTATTTT AGCTGGTATC GGTGTCTTCC	480
TTGCCTATGT AGGGATTAAG AATGCTGGAC TTTTGAAATT CACGATTGAT CCAGGCAACT	540
ATACTGTGT AGGAGAAGGG GCTGACAAAG CTCAAGCAAC GATTGCAGCA AACTCTTCAG	600
CAGTTCAGG ATTGGTCAGC TTTAATAATC CAGCTGTTTT AGTGGCTCTT GCAGGACTTG	660
CCATTACTAT CTTCTTTGTC ATC	683

## (2) INFORMATION FOR SEQ ID NO: 339:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 852 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

CTACTTTACA TGAAGTAGT CACTGAATTC CAGTTAGAAA TTACTTTGTA ACTACGTTTT	60
GAGGAGGAGT AAAATGCTTT CCTACGTTTCG ATATTACCCA CTAGCGATAG CTAATTAAT	120
GTGTCTGTGC TCTCCTAAAA TCTGCTGATT TATTACTGAC TAATACAGGA GGTTTTTTTT	180
ATGgACAGAC AATCATATCT GCTATTGGTG TTTATATTTT CACCAGTATC GATTATTAA	240
TTATTTTAAT TATTTTATTT GCACAGCTAT CACAGAATAA ACAGAAATGG CATATTTATG	300
CGGGGCAATA TCTAGGCACA GGCTTACTTG TAGGGGCGAG TTTAGTTGCT GCTTATGTCG	360
TTAATTTCTG GCCTGAAGAA TGGATGGTTG GATTGCTTGG TTTAATCCCT ATCTATTTAG	420
GGATTCGCTT TGCAATTGTT GGAGAAGATG CGGAAGAAGA AGAGGAAGAA ATTATTGAAA	480
GATTAGAACA AAGCAAGGCA AATCAACTGT TTTGGACAGT TACATTGCTG ACAATTGCGT	540
CTGGCGGAGA TAATTTAGGT ATCTATATAC CTTATTTTGC TTCGTTAGAT TGGTCACAGA	600
CCCTCGTGGC CTTGCTTGTG TTTGTAATCG GCATAATTAT CTTTTCGAG ATTAGTCGGG	660
TGTTATCCTC TATTCGGTTA ATATTCGAGA CAATTGAAAA ATACGAGCGA ATCATTGTGC	720
CCTTAGTATT CATTCTACTT GGACTATACA TCATGTATGA AAATGGCAGC ATAGAGACTT	780



1361

TTCTGATCGT GTAGATTTTT TTGTTTCACT AGGGATTAG CCCGAGCTCA AATCAGCTCT 840  
 CTGATTTTCA GA 852

## (2) INFORMATION FOR SEQ ID NO: 340:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 754 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

CCGCACAAAA GCGCATAGTA TCAAGATTCT ATAAAGCCTT GATACTATGC CTTTTTAATG 60  
 GATAAATAGT TAGTCTTTT TAAAGACCGG ATCTTTCAAA CTCTGCATAC TGGCATTGAT 120  
 CACCGCGCCT AGGATAACAA TTTAGCAAT CAAGATAAAC CAAAACATCA TAACAACAAG 180  
 AAGAACGGAA CCTAAATTC GGACATCCAC CAAATGATGG ACATAGTAAT TGAGATAACT 240  
 AGAGAACAGA GTTAGTAAAC CTAAATCAC TAAGAGAACA AAGGCACTGC CTGGTAGGGT 300  
 ATAGCTAATT TTCCTGTTAG ATAGATTGGG AAGAAAATAA TAAAGCATGA CCAAGATACC 360  
 AAAGAGGAGG GCGTAAATCA GAGGACCTGC CAACCTTGT AAAGCCTGAT AGATAATGCC 420  
 ATCTTTTGTG CAATAATGAG CAAGTAAAGC CAAATCATC TGACCAAATA AGATCAAAAA 480  
 CAAGGCAAAC GCAAAGAGGA GCTGCAACCA AACTGACTA GGAGACTTAG CATCTGATGG 540  
 GAAATAAGTC CACGACTCTT TTCGACGCCA TAAGCCTTGT TAAAAGCTTT TTGCAAGAAA 600  
 TTCATAGATT TTGAAAAACT CCATAACGCC GATAAACAG AAAAAGCTCA TAAACCTGTT 660  
 GAAGGTTGCG TCAAGACTTC TCTGGCTATT TTTCCACAC CTTCATAGAG GCTTGGGGGG 720  
 CAGACGTCTT TCATAAGGCC CAAAATTCT CCCA 754

## (2) INFORMATION FOR SEQ ID NO: 341:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 707 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

GGGGATAACT CTAGGAGTAC CGCTATTACT CGACTTAATG AGTGCACAAG AAGTCAGGAT 60  
 TTTTATGCAG GTTGGGCGCT TCATCAGACA GGGAAGATTT ACAGCGACTA TTATGGAAGT 120

1362

CAAGGTTTGC TTTATTATTT GCTGACTTAC GTGAGTCAGG GCGGATTTTT CTTTGCCATC	180
TTTGAGTGGT TAGCCTTGGT AGCAGGAGGA TTTTTCCTTT TTAGATCAGC GGACACCTTG	240
ACAGAGCAAG GAGACCAAGC TGGACATCTG GTGACTATTT TTTACATGCT AGTTACAGGT	300
CTTGCTTTTG GTGGAGGCTA TCGGACTCTT TTAGCGCTTC CTTTCTTATT CGCAGCCTTT	360
AGTTTAGTTG CGGCTTACCT AAGCAATCCA AGCCATGATA AGGGATTGTG ACGGATTGGG	420
CTAGCTTTGG CAGGCGGATT TTTCTTTGCT CCCTTATCAT CGCTCCTGTT TATTGCTGTA	480
GTGAGTTTAG GCTTGTGGT CTTTAACCTT GGGCATAGAC GCTTTGCGCA TGGGTTTTAT	540
CAGTTTCTTG CAGTGGCTTT AGGTTTTTCA CTGTCTTTT ATCCAATGC CTACTATAGT	600
GCTGCAACAG GAAGTTTTGG GGATGCGWTT AGTGGTATTC GTTATCCTAT TGACAGTATT	660
CGCTTTGATT TTAATTCTAA AATTTTAGAG AATATGTTTT TTAAAGG	707

(2) INFORMATION FOR SEQ ID NO: 342:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 762 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

GGATTTTGAA AAACCATACC GATTTGACGA CGTATATTCC AAACATTTTC CTCAGTCAAA	60
CGTTGGCCAT CAATTACAAT CTCTCCGGAT TCTGCTTCCA GTAAGCCATC AATTAATCGA	120
ACCGTCGTTG ATTTACCACT ACCATTATGC CCTACAATCG AAAGCCATTC TCCACGTTTC	180
ACGTGAAAgT AATATCCTTC ACATCGTAGT AGTTCTGATT TTCTTTATAG CGAAAAGAAA	240
GATTTTTTAC ATCAATTATT GATTTCAATT CGAACCAAT GTCCCTTTAA ATACATAGGC	300
ACTACCCTTG AAATAGTCAT AGCCAGAGTA GATAGTGAAA AATAAGGCTA CATAAAGTAG	360
AACTTGACCA AGCAAAGTCC AATGTAATAG CAAGAAAATA ATGGCAAACA TCTGACTAAA	420
AGTTTTAATT TTTCCAGGCA TTGCTGCTGC TAAAATTGTT CCACCAGTTT CAACCAATAA	480
AAGCCTTAAA CCTGTCACAG CTAATCAGC ACAGATAATC ACTGCAACAA TCCAAGCCGG	540
AGCCATACCT AACTCAATCA ACATAATAAA AGCCGACATA ACTAGTAACT TATCCGCCAT	600
AGGATCTGCA AATTTACCAA AATTACTGAC CACATTCCAT TTACGAGCTA AATATCCATC	660
TAAATAGTCG GTAATACTGG CAACAGCAAA GATAATAGCT GCAACTATAT GACTCTCTAT	720
CGAATTTCTT ATCGTTAAAA TAAAGATAAA AATAGGTATA AA	762

(2) INFORMATION FOR SEQ ID NO: 343:

1363

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 482 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

```

CTTTTGATAC ACTTAAACTA TGAATACAAA TCTCAAGCCC AAAC TTCAGC GTTTTGCTTC      60
TGGGACTGCC TTGCCTGTC CTATCTGTCA AGAAAATCTG ACTCTGTTAG AGACTAATTT      120
CAAGTGCTGC AACCGTCATT CTTTGACTT GCGGAAATTT GGCTATGTCA ATCTAGTCCC      180
TCAAATCAAG CAATCTGCTA ACTACGACAA GGAAAATTTT CAAAACCGTC AACAAATCCT      240
AGAAGCCGGC TTTTACCAAG CTATCTTAGA TGCTGTATCT GACTTGCTTG CAAGCTCAAA      300
AACTACCACA ACAATTTTGG ATATCGGTTG TGGTGAAGGA TTCTATTCTC GCAAAC TACA      360
AGAAAGTCAC TCTGAAAAAA CTTTCTATGC CTTTGACATC TCCAAAGATT CAGTCCAAAT      420
CGCGGCTAAA AGTGAACCCA ACTGGGCAGT CAATTGGTTC GTTGGCGACT TGGCAGCACT      480
TC                                                                                   482

```

(2) INFORMATION FOR SEQ ID NO: 344:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 520 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

```

TTTATTTTAA TAAAGTCAAT ACCTGTCTTT ACTTTTCTT AAAAAAAGTT TATTATGTTC      60
TTTAAGGAGG TGTAACAT GAAAATAAAT AATAAACTCG TTGGAGAACG TATTCAAAT      120
ATCCGTTTAA GCCATGGCGA CTCTATGGAA AAATTTGGAG AAAAATTTAA TACTAGCAAA      180
GGTACAGTTA ACAACTGGGA AAAAGGTCGC AATTTACCAA ATAAAGAAAA CCTACTAAAA      240
ATTGCATCTA TTGGAAAAAT GAGTGTGAA GAGTTACTCT ACGGCGATTA CAATACTTAT      300
CTACACTTAA AGATTATGGA TTTAGCTCCT GAATGTATAA AAAATTATGA TGAGTATAAC      360
TCTTTACACG ATGATATAAC AAATAAAGCG TTACAGATCG CTCAAAATAC CATTTCTAAG      420
ATTGATTATC AAATTTTACA CGAAACGATC AAAAAATTTA TTGATTTAGC TATCGAACAA      480
TCGAGAGATT TGCAAGGAAA TTTGTTGAAA AATAACGGGT                                520

```

1364

## (2) INFORMATION FOR SEQ ID NO: 345:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:

```
GCATCAAATC CGCCATCAAA GAAGTTCTCT GGATTACCA AGACCAGTCA AATAGCTTAG      60
AAGTGCTTAA TGACAAGTAC AATGTTCACT ACTGGAATGA CTGGGAAGTT GGAGACACGG      120
GAACCATTTG TGAGCGCTAT GGTGCCGTTG TTAAGAAACA CGACATTATC AATAAGCTTC      180
TCAAACAGTT GGAACCAAT CCTTGAACC GCCGCAATAT TATTTGCTC TGGGATTACC      240
AAGCTTTCGA AGAACAGAT GGGCTGCTCC CGTGCGCCTT TCAGACCATG TTTGATGTTT      300
GGCGTGTTGA TGGGGAATC TATCTGGATG CGACCTTGAC CCAGCGCTCC AATGATATGC      360
TGGTGGCCCA CCACATCAAC GCTATGCAGT ATGTGGCTTT GCAGATGATG ATTGCCAAAC      420
ATTTTGCTG GAAGGTTGGG AAGTTCTTCT ACTTCATCAA CAACCTCCAT ATCTATGATA      480
ATCAATTTGA ACAAGCTCAG GAATTGCTCC GTCGGGAGCC GTCAAAGTGC CAACCACGCT      540
TGGTTTTAAA TGTTCCTGAT GGGACTAATT TCTTTGATAT CAAAGCAGAA GATTTTGAGT      600
TGGTGGATTA TGACCCTGTT AAGCCACAGT TGAAGTTTGA CCTAGCTATT TAAAAGAATA      660
GAAAAAAGAA GTTGAGAATA ATCCCAACTT CTTTGTGTTT TTAACGTGAT ACGCGGCGAC      720
GAGCTGCTTT TTTACGGTTT TCTTCGATGA AAGCTGCTTT TTGCTCTTCT GGTTCGATTA      780
CTTTCTTTT AAATGCGTAT ACTGCACCTG CAACGGCAGC GACAGTTCCT GCGACACCTG      840
TTACAAGACC TTTAGCGAAT CCTTAGCCA TGAGTCTTCC TCCTTTATAT TCTCAATCAG      900
CCAGCCTCCT CAAGAGGTCA CATTTCCTG ACTGACCTTT TTGTGTTATA ATAATAGTAA      960
CGAAAAATG GGAATTTTTC AAGGAAAAAA GATGAGAACA AAA                        1003
```

## (2) INFORMATION FOR SEQ ID NO: 346:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 750 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

```
CCGCACGTAC TATTCCAGAT GCCGAGGAAG TGGACCTCAT CCTCGTTGGC GCAACTGGTC      60
```

1365

TCAACGCCTT TGAACGCCTC TTGGTCGGCT CTTTCATCTGA ATACATACTC CGCCATGCTA	120
AGGTCGATTT GCTGGTTGTG AGAGAACAAG AAAAAACCTT ATAATCACAA AGAAAAGGAG	180
CCCCTAGCTC CTTTTTGTTC ACGATTATT TCTCTCTTTA TGGCGTTCGT AAGCCTTGAG	240
CTGGCGCTGC AGTTCCTTTT TAATAGCAGG TTCTGGAGCA TATTTTCTT CCCAATTATC	300
TGGTTTAAAG ATTTTATGGG TCACTGGATC AAAATGAGCC TTGCCATCTG GAAAAATTTT	360
CCCCATATTG GCCTGATGGA CAATATCAAA AATACGTTCT GGGTCCACCC CCATCAAGAC	420
AAAACGCGG TAGGTGAAGT AAAGCGTGTC AATCAAGGCA TCCACTTGCC CTATCAAATC	480
TTGCTGAGCA GGTGTCTTCT TGGCTACTTT ATCTGCTGCC TTATCAAGGG CCTGATGAAG	540
TTGGGATACA GCTTGACCAA AATCTTCTTC AGAAGGACTG GCTGCTCGAA CAACTCCAC	600
CAATTCTTCT ATTTTAAAC CAGCCCTATG GGTTCACCC TCTAAATCCC AAGCTCGAGG	660
TTCTTCTTGG GTTCGTTTAT CCATCATGTG GTGAAAAGTC TTGACCTTAT TGAAATGATA	720
GTACGGCTG ACAAAGACTT TTTCTGAAGA	750

(2) INFORMATION FOR SEQ ID NO: 347:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 596 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

CGCAACATAC GGATAACCTC CAAAGAATAT TTTTATATTA TAGCAAAGCT TTAAATTGAA	60
TGTTAGAGTC TTGTTCAAAA CAATCATCAA AACCACGTGG ATGATGGTAT TCTACTAAGT	120
GTTGATCTTG AGGATAAGTG TACTTACCGC CAACTTCCCA GATAAATGGA TGGAAATCGT	180
ATTGCAAGCG ATCTTTTCGC ATTTTCAAAA GTTCTAGAAT CTCATTAGTA GAAGCCATGA	240
AGTTAGACCA GATATCATAG TGAAGTGGGA TAATGACTTT GGTACGCAGA TTTTCTGCCA	300
TACGAAGAAG GTCGATAGAT GTCATTTTGT CTTGGATACC TACCGGATTT TCACCATAGT	360
TATTCAAAGC AACATCAATT TTAAAGTCTT TACCATGTTT TGCAAAATAG TTTGAGAAGT	420
GAGAATCTGC ACCATGATAG ATGGTTCCAC CTGGTGTTC AAAGATATAG TTAACAGCCT	480
TTTGAGCCAT TTCTTCATCT GTAACAGCCA AGCCAGCAGT TCACCGCCTG TCTCATCAGC	540
ACCGTTCACT GGGAGAGTTA CCAAGCAAGT ACGGTCAAAT GATTCTACTG CATGAA	596

(2) INFORMATION FOR SEQ ID NO: 348:

1366

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 673 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

CAGAGTCAAC AGCCTGAGTT GAAGGCAACT TTAGACACAG CAGTTACGAC AGCTGAATGA	60
GCTCCTCCAT CAGTTTTTTC TTTAATGAGT CCAGCTACAT CTTCAACTTC GAGGCCGTTA	120
ATCACAATGT CAGCGCTAC TTCTTTTGCA AGGGCAAGTT TGTCAATTGT GATATCGACT	180
GCGATAACAT GAGCATTGAA TACTTTTTTA GCGTATTGAA CAGCGAGGTT ACCAAGTCCA	240
CCAGCACCGT AAAGAACAAC CCATTGGCCT GGTTCAACTT TTGCTTCTTT GATAGCTTTA	300
TAGGTTGTTA CTCCAGCACA TGTGATAGAA GAAGCTTGGG CTGGATCAAG TCCGTCAGGA	360
ACTTTGACAG CATAGTCAGC AGTTACGATA CATTGTTTCAG CCATACCACC GTCTACTGAG	420
TAGCCAGCAT TTTTCACTGT ACGGCAAAGG GTTTCGCGAC CAGTTGTACA GTATTGCGAA	480
GTGCCACATC CTTCAAAGAA CCAAGCAACG CTGACGCGGT CACCGACTTT AAGGCTTTTC	540
ACATCTGGAG CAATCTCTTT AACGATACCG ATACCTTCGT GCCCAAGAAC ACGTCCTGGG	600
ACTTGACCAA AGTCACCATG AGCAACGTGG AGGTCGGTGT GGCAAACGCC CACAGTATTC	660
ACTTCTACAA GTG	673

(2) INFORMATION FOR SEQ ID NO: 349:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 198 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

GTACCCTACA AATGCTTTAC AGTATGGGTT GAGGGTGGTC AATGGAAC TAAGTAGGT	60
TGGACAGGAA CTTTGGATA TTCTGATTAC TTACATTCTA CTCGATATCA TACAGCAACT	120
GTTAGACATG GGGGTAGAAC CTCTAAGGAT TATGCAAAAC CTGAGGCATG GGCTAGAGCT	180
TCCCTCACCA AGATTCCG	198

(2) INFORMATION FOR SEQ ID NO: 350:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 891 base pairs  
 (B) TYPE: nucleic acid

1367

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

GCTTCTTCTA TAGACAAAAA TATCATGGGT AAAATAATCA AGGCTATAGC TAGAAGGAGG	60
GACCAATCCA CTAATAATCC TAAGAACAAA ACACTCAAGA GAGCAGAAGA GAGAGGTTCA	120
CTGGCACTGA TAACGGCAAC CACCAAAGGA GAAACCAAGG ACACAGCCTT CATGGAAATG	180
AAAAAAGCAA AAGCCGTTCC AAAGAAAGCG ATAATGAGGC AAATCAAGAT ACTCCAAATA	240
TCAAGAGTAA AGGAAAGCTG ATAAACCGGC GAGAGGACAT TGCTAAACAA ACCTGCCAAA	300
ATCATCCCCC ACCCAACCGT AGGAACAAAA CCATAACGCT TAGCAAAAGG TTGGGGCAAG	360
ATAACATTAA ACATAACACC CATGGCACTC AGCAAACCTG TTATAAGAGC TAGCGGCGTC	420
ATGGATAACT GAGAGAGGTC TCCCTTTGTC GCCATCAAGC AAACACCCAG CATGGCAACC	480
AAAACATAGA AAACAGCGCT TTTTGACGCT CGTTTTTGAT AAACCAAGCG ATTGTAAAG	540
AGGATAAAGA CAGGGCTAAT AAAGTGAAA ATAGTTGCTG TCGTAGCATT TGAGTATTCT	600
ACACAGAGAT AGAAAAATA CTGAACTGAA AAAATCCCCA AAATAGCATA GGCTAAAAAG	660
GGCAGGTAAT TTTTCTTGTC TCGCCAAATA TCTAGCACTT GCGATTTTAA TTGTATTGCA	720
GACCAATGA GTACAAGACT CCCTGCCAGT GTCAAACGCA TAGAGGTAAT CCAGCCCGAA	780
GACACCTGAT AATGAGTAAA GAAGTACTCT CCTAAAATTC CACAGATTCC CCATATTAAG	840
CCGGATAGGA GCGAATAAAT TTTTCCGTTA ACAATCTTTT TCTGATACTG A	891

## (2) INFORMATION FOR SEQ ID NO: 351:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 325 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

GAAAGCGTTC AATAGAACAT TGCTTTTTTA TTTTAGAGT AAGCTAAGCG CTTGAGCATC	60
TGCGATGATG GTTACATCAG GGTGATTTTG GAGGCTACTT GCAGGTAGGT TCTCAGTCAC	120
TGGGCCAGAT ACTGTTCCGG CAATGGCTTC TGCTTTTCGAC TCACCGTAAG CAAAAAGAAT	180
AATAGACTTG GCATCCAAAA TGTTTTTAAT CCCCATTGAA ATAGCTTGGG TTGGGACGTC	240
TTCAATCTTG GCAAAGAAGC GTGCATTGGC TTCGATAGTA GACTGGTCAA GTTCTACTAG	300

1368

ATGCGTTTGA CTGTCAAATG GAGTG

325

(2) INFORMATION FOR SEQ ID NO: 352:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 344 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

CAAGAGCAGT TTGATGATTT TTGATAACCA TCGGAATTTA AAATACAAAT ATGGCAATCG	60
CAAGTTTTGG TGTAGAGGCT ATTATGTAGA TACGGTAGGC CGTAATCAGA AAGTGATAGC	120
TGAATATATT CAGAATCAAT TACAAGAAGA CAGAGTAGCA GACCTAGCTC ACGTTATTGC	180
AGTCAGTAGA TCCGTTTACT GGCGAAATAA ATAAGAGGAA GTAACGTAA GTGCTTTAGC	240
ACCTGCTCGG GAAAGTGGTG CGGAGGAAG CTATTTTCAGG ATGCTTTGGC CCTGGCCGGT	300
AGAAGCGTTA TAGCCGCAGA CTACGACACT TCACACTGGT GGTT	344

(2) INFORMATION FOR SEQ ID NO: 353:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 692 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

CCCTATCCCT GCTATTGGGG CTGCTCTCAT TGCTGCTTG GCACAAATCA GTCTTCCAAT	60
TGGACCTGTT CCCTTCACTC TGCAAAACTT TGCAATCGGC TTGATTCTAC TGTCTTTAGA	120
CCGAGAGAGG CTGTACTTTC TGCTGGACTC TATCTTCTTC TAGGTGCTAT CGGTCTTCCT	180
GTCTTTGCAG GAGGTGGAGC TGGTTTTCAG GCTTTAGTTG GCCCTACTGC AGGCTATCTT	240
TGGTTTTATC TCGTTTACTC TGGACTTACT TCCTCTCTAA CCAACAGCAA GAGTGGTGTT	300
GTAAAGATTT TTCTTGCAAA CCTCTTGGGT GATGCCCTTG TCTTTGTCGG CGGGATTCTC	360
AGCTTGCAAT TCCTAGCTGG AATGGCATTT GAAAAAGCTC TTGCTGTGGG GGTTCCTCCC	420
TTTATCATTC CAGACCTTGG CAAACTTCTA GCTATTAGTT TTATTAGCCG TCCCCTACTT	480
CAACGCCTTA AAAATCAGGC TTACTTTACT AACTAAAAAA GGATATCGAG TTATCATGAC	540
TCAATATCCT TTCTTTTAT TTTGAAAAC TATACTCAAT GAAAATCAAA GAGCAAAC TA	600
GGAAGCTAGC CGCAGGCTTG CAAAACACTG TTTTGAGGTT GTGGATGAAA CTGACGAGTA	660



1369

AnATCTCATA CATACGGCAA GGCAAAGCTG AC

692

## (2) INFORMATION FOR SEQ ID NO: 354:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1005 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

GTGATGGACT ACTGGTTCAA AACGCATCCA GAAGATTTT TCGATAATGT CGGACCTCTT	60
GTAGCCAGTA ACTTTTTTCA TACTTACACC GAAGATTTC ACTTGATGAA GGAAATGGA	120
GTTAATTCTT TCCGCACTTC CATCCAATGG AGTCGACTCA TCAAGAATT AGAGACAGGT	180
GAGCCTGATC CAAAAGGTAT TGCTTTCTAC AATGCCATCA TTGAAGAAGC TAAAAAGAAC	240
CAGATGGATC TTGTGATGAA TTTACATCAT TTGATTTC CAGTGGAAC TCTTCAAAAA	300
TACGGTGGTT GGGAAAGCAA ACATGTAGTG GAGTTATTCG TGAAGTTGC CAAGACTGCT	360
TTACATGCT TTGGAGATAA GGTTCATTAC TGGACAAC TTCAATGAGCC AATGGTCATT	420
CCAGAAGCAG GGTACTTATA TGCTTTCCAT TATCCAAATC TAAAAGGAAA GGGAAAAGAG	480
GCCGTACAAG TCATCTATAA TCTAAACCTT GCTAGTGCAA AAGTGATTCA ACTATATCGC	540
TCATTAGAAC TTGATGGAAA GATTGGGATT ATTTTAACT TGACACCTGC TTATCCAAGA	600
AGTAATTCTC CAGAAGACTT AGAAGCAAGT CGATTTACAG ATGACTTCTT TAACAAAGTC	660
TTCTTGAATC CAGCTGTTAA AGGAACCTTC CCAGAAAGAT TGGTAAAACA GCTAGAGAGA	720
GATGGCGTGT TATGGAGTCA TACCGAAAAA GAGCTTCAAC TGATGAAATC AAATACGGTT	780
GATTTTCTTG GAGTAACTA CTACCATCCA AAACGTGTT AAGCACAAGC AAATCCTGAG	840
GAATATCAGA CGCCCTGGAT GCCAGACCAA TACTTCAAAG AGTATGAATG GCTGGAGCGT	900
CGCATGAATC CATATCGTGG TTGGGAAATT TTCCGAAAG CCATTTATGA TATTGCTATG	960
ATTGTGAAGG AAGAATATGG TAATATCCCA TGGTTTATCA GTGAA	1005

## (2) INFORMATION FOR SEQ ID NO: 355:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 973 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

1370

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

```

CCGACAAGCA ATATTAAAA GAGTAAACTA TTAAGTAGTT AATTAACCGG TTTATTACTT      60
TATAGTGAAT CAAATATACT TAAGAAAAGA GGAAAGAATG AAAATTAATA AAAAATATCT      120
AGCAGGTTCA GTGGCAGTCC TTGCCCTAAG TGTTTGTTC TATGAGCTTG GACGTTACCA      180
AGCTGGTCAG GATAAGAAAG AGTCTAATCG AGTTGCTTAT ATAGATGGTG ATCAGGCTGG      240
TCAAAAGGCA GAAAACTTGA CACCAGATGA AGTCAGTAAG AGGGAGGGGA TCAACGCCGA      300
ACAAATTGTT ATCAAGATTA CGGATCAAGG TTATGTGACC TCTCATGGAG ACCATTATCA      360
TTACTATAAT GGCAAGGTTT CTTATGATGC CATCATCAGT GAAGAGCTCC TCATGAAAGA      420
TCCGAATTAT CAGTTGAAGG ATTCAGACAT TGTCAATGAA ATCAAGGGTG GTTATGTCAT      480
TAAGGTAAAC GGTAAATACT ATGTTTACCT TAAGGATGCA GCTCATGCGG ATAATATTCTG      540
GACAAAAGAA GAGATTAAAC GTCAGAAGCA GGAACGCAGT CATAATCATA ACTCAAGAGC      600
AGATAATGCT GTTGCTGCAG CCAGAGCCCA AGGACGTTAT ACAACGGATG ATGGGTATAT      660
CTTCAATGCA TCTGATATCA TTGAGGACAC GGGTGATGCT TATATCGTTC CTCACGGCGA      720
CCATTACCAT TACATTCCTA AGAATGAGTT ATCAGCTAGC GAGTTAGCTG CTGCAGAAGC      780
CTATTGGAAT GGGAAAGCAGG GATCTCGTCC TTCTTCAAGT TCTAGTTATA ATGCAAATCC      840
AGCTCAACCA AGATTGTCAG AGAACCACAA TCTGACTGTC ACTCCAACCT ATCATCAAAA      900
TCAAGGGGGA AACATTTCAA GCCTTTTACG TGAATTGTAT GCTAACCCCT ATCAGAACGC      960
CATGTGGGAT CTG                                          973

```

(2) INFORMATION FOR SEQ ID NO: 356:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 843 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

```

GGTCGCATCT GCAATATCTG TCGCCTCCAC ATAAGCGACA CCAGCCTTGT CTGCTGCCCG      60
TTTGACACGT TCTGCAGATT GACCCAGGAT GACCATCTTC TTGAGTCCAG TAATGTCTGG      120
CACCAATTCTG TCAAATCAT TGCCACGGTC CAAACCACCT GCAATCAAGA CGACCTTGCT      180
GTTGTCAAAT CCTGACAAGC TTTTGTAGTA GCCAAGATAT TAGTTGATTT ACTGTCGTTA      240
TAGAATTTAA CACsCTTGAT GTCATCCACA AACTGGAGAC GGTGTTTGAC ACCACCGAAG      300
GCTGAAAGAG TTTCTTGAT GGTGTTGATTG TCCACATCAC GAAGCTTGGC TACAGCAATA      360

```

1371

GTCGCAAGGG CATTTTCCAC ATTGTGGCTA CCTGGAACAC CGATTTTCATT CGCTGCCATG	420
ACTACTTCAC CACGGAAGTA GAGTTGACCA TCTTCCAGAT AAGCTCCATC AACCTTTTCA	480
AGTGTTGAAA ATGGTACAAC AGTGGCTTCT GTCTTGGAAG TCAAGTCTTT TGCCAAGTCT	540
TGATTAAAGT TCAAGACAAG GAAATCAGCT GCTGTCATCT TGTTCGGAT ATTCCACTTG	600
GCTGCTACAT ATTCCGAAAA TGACCCATGG TAGTCGATAT GAGTTGGCAT GAGGTTGGTA	660
ATAACCGCAA TCTCTGGATG GAATTCCTTGA ACACCCATGA GTTGGAAGA AGAAAGTTCC	720
ATAACAAGCG TGTCTTATC TGATGCTATT TGAGCAACCT GACTAGCTGG ATAGCCGATA	780
TTCCCTGATA AAAGACCATG TTGGCCAGCA GCAGTCAAAA CTTCCCGGGn TCCTCTAGAG	840
TCG	843

## (2) INFORMATION FOR SEQ ID NO: 357:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 807 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

TTTTTTTAT ATTTTTTTTA TTTATTATTT TTTGGCAAAA AAGACCAATT TGCTTTGGAG	60
CATTGCTTCT GCATTAAATT GTCTATTTT GTCGTGCTG TTACGCTCTT TGTATCATGT	120
ATTAAGTAGC AAGTGCAACT TGCAAACTAC TAGTAAGAGG AGAAAAACAA AATGGTTATG	180
ACTGACCCAA TCGCAGACTT CCTAACTCGT ATTCGTAATG CTAACCAAGC TAAACACGAA	240
GTACTTGAAG TACCTGCATC AAACATCAAA AAAGGGATTG CTGAAATCCT TAAACGCGAA	300
GGTTTTGTAA AAAACGTTGA AATCATTGAA GATGACAAAC AAGGCGTCAT CCGTGTATTT	360
CTTAAATACG GACCAAATGG TGAGAAAGTT ATCACTAACT TGAAACGTGT TTCTAAACCA	420
GGACTTCGTG TCTACAAAAA ACGTGAAGAC CTTCCAAAAG TTCTTAACGG ACTTGGAATT	480
GCCATCCTTT CAACTTCTGA AGGTTTGCTT ACTGATAAAG AAGCACGCCA AAAGAATGTT	540
GGTGGTGAGG TTATCGCTTA CGTTTGGTAA AATCAAGATA CAAAGCTCGT AAAGAACAAA	600
GCAAAATTAG GAAGTTGGAG AAGTTTGTTC ACAACAGGC CAACTTATCT ATTTTGACAA	660
GTTCTTAGAG CGTGTTCACT TCAGCTCTTG AGCTAAGTAA GTATCTGAAC CCCGTGAAAA	720
CTGGCCGTGC TGGCATGTTT GGGTAACAGG AGAATAATAA CATGTCACGT ATTGGTAATA	780
AGTTCAGCTA AGGCCTTCGT AAAAGTT	807

1372

## (2) INFORMATION FOR SEQ ID NO: 358:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 653 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

CCCAGTATTT TTGTCCAAGC ACGACCAGAA AAGGATGATA CAGATCTGGA ATTGGCTCTC	60
TTAACCATCT tTGAACAAAA TCCTCAGGCT CAGGTCACTA TTTTCGGTGC CTTGGGTGGC	120
CGTATTGACC ATATGTTGGC CAATGTCTTT CTGCC TAGCA ATCCTAAGTT GGCACCTAT	180
ATGCATCAAA TAGAAATTGA GGATGGGCAA AACTTGATTA CTTATTGTCC AGAAGGAATC	240
AGTCAGCTAG AACCTCGTTC AGACTACGAC TATCTAGCCT TTATGCCAGT TCGGGATAGC	300
CAAGTATGAG TTGACAGAGG AAAATTTTTT CTTTAAAAAA GTGTACGCTT CTAACGAATA	360
TATAGATAGG GAAGTGTCGG TAACTTGCCC AGATGGTTAT GTGGTCGTAC TGCATAGCAA	420
GGACAGGAGG TAGGATGGAA AGTTTACTTA TTCTATTATT AATTGCCAAT CTAGCTGGTC	480
TCTTTCTGAT TTGGCAAAGG CAGGATAGGC AGGAGAAACA CTTAAGTAAG AGCTTGAGG	540
ATCAGGCAGA TCATTGTCA GACCAGCTGG ATTACCGCTT TGACCAAGCC AGACAAGCCA	600
GCCAGTTAGA CCAAAAAGAT TTGGAAGTGG TTGTCAGCGA CCGTTTGCAA GAA	653

## (2) INFORMATION FOR SEQ ID NO: 359:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 641 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

CACCATGTGA TGTGACGCTG GCCACAGCTG TCAGAAATCT GGCGAGCCAT CGTGTGCAAT	60
GACTCTTCCC GATGTAATCT TGTTCATAGT CCTTTGATGA ATATGTTCAA GCTGTAGAAG	120
GTGCGCTTCC TGAACACTTA TCAACTGTTA CAGGCGAGTT GACCAGTCAG GAAACAGATG	180
GCTGGTACAC ACTTGCCAAC ACTTCTTCAT CCCGCATTTA CCTAAAACAA GCCTTCCAAG	240
AAAATAGCAA CCTCCTAGAG CAAGTGGTAG AACCTTGAC TATTATCACT GGTGGACACA	300
ACCACAAGGA CCAGTTGACC TATGCTTGGG AAACACTTTT GCAGAATGCG CCACATGATA	360
GTATCTGTGG CTGTAGCGTG GACGAAGTTC ACCGCGAGAT GGAACGCGT TTTGCCAAGG	420

1373

TCAACCAAGT AGGAAACTTT GTTAAAAGTA ACTTGCTCAA CGAGTGAAG GGTAATTTG	480
CTACGGATAA GGCTCAAAGT GACTATCTCT TTAAGTGCAT TAACACAGGC TTGCATGATA	540
AGGTCGATAC TGTCAGCACA GTGATTGATG TGGCGACTTG TGATTTCAG GAATTGCACC	600
CAACAGAAGG CTACAAAAG ATGGCTGCTC TTATCTTGCC G	641

## (2) INFORMATION FOR SEQ ID NO: 360:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1958 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

CCTCAAGGCC AATTTGAAGG CTCTAAAACA ATGGAAAAGT GCTACACAGA TGTGACAGAA	60
TTTGCCATTC CAGCAAGTAC TCAAAAGCTT TACTTATCAC CAGTTTGTAGA TGGCTTTAAT	120
AGCGAAATTA TTGCTTTTAA TCTTTCGACT TCACCCAACT TAGAACAAGT ACAACAATG	180
TTAGAACAGG CATTCAAAGA GAAGCACTAC GAGAATACGA TTCTCCATAG TGACCAAGGC	240
TGGCAATATC AACACGATTC TTATCATCGG TTCCTAGAGA GTAAGGGAAT TCAAGCATCT	300
ATGTCACGCA AGGGCAACAG CCAAGACAAC GGTATGATGG AATCTTTCTT TGGCATTTTA	360
AAATCCGAAA TGTTTTATGG CTATGAGAAA ACATTTAAAT CACTTAACCA ATTGGAACAA	420
GCCATTATAG ACTATATTGA TTACTACAAC AACAAACGAA TTAAGGTAAA ACTAAAAGGA	480
CTTAGTCCTG TGCAGTACAG AACTAAATCC TTTGGATAAA TTAATTGTCT AACTTTTGG	540
GGTCAGTACA AAACTCTTGC TACTATGCGT TTTATTATTG AAAGACTTAT TGGACTTTCT	600
CTCAAATCGA GTTTTTACTC AATTTTCTTA CTTGATTGGG ATTGAAATTC CAATTAATTT	660
CTCTGAGTAG AGTGTCTTGA TATTGGCTTC ATCAACAGAG GCCTTATCAA TTTTACGTTT	720
CAAGAAAAAT TCTTGAATGG TTTGATTTC AGGCTCACGA ATAGCACGGT GTTGTGTTGA	780
GATGAGGATT TCATAGTGAA GCGGAGCTTG GGTAAAAATA ACATCTGTAT TCCCTGCAGA	840
ATAAACCTCA ACAAGGTTG CATCGGTACT TTCTAGCTGA CTTTTTACAA GTTGCGAGTG	900
TGAGTTTGTC GTATTGATAA GCTTCATAAT ATTTCCTCCG ATTTTCTAAT TCTATTATAG	960
CACTTTTGA ATAAAGTCGC TTGATTTATA CTCAATGAAA ATCAAAGAGC AACTAGGAA	1020
GCTAGCCGCA GGCTATACTT GAGTACGGTA AGGCGACGCT GACGTGGTTT GAATTTTATT	1080
TTCGAAGAGT ATTAGCCAAT CTTATGCTGT TTTTCCAAG ATTCAATGGC CCATTTATGG	1140

1374

CTACCACGTT TAAGGTTTTT GATAGCCTCG TCAATAGGGA ACCAGGCAAT ATGATTAAAG	1200
TTTTCTAGTG GCTTTTGTTT TTCTTTGAAA GGAGTTGCTT CATAGAGGTA GGCAGGATTG	1260
TAGTAGTAGG TATCACGATG ACGAGAATAG AAATATTCGT CAGCTTGTCC GTAATAGGTA	1320
CCAATTTCTG CTGTGAAACC AAGCTCTTCA ATCAACTCAT GCTTTAGGGC TTCCTGATGA	1380
TTTTCACCTG CTTCAATTTT TCCACATGGT AGGAACCAAG CACCATTGGG TTCTTGAACA	1440
AGAACAATTT GTTTTGTTC AGGATTAGGG ATAATGCGT ATACGCCATA GCGAGCAATA	1500
TAGTCTGTAT TCACTTTTTT TCTCCGAAAG TTGGGTTTGC CATTCGATTT TCCTCATTAT	1560
CTAGTATCGT TATTATTATA GTGAAATGAA CCAAAAATAG TACACAATGT GGTATAATCT	1620
TCTTATGGCA TATTCAATAG ATTTTCGTAA AAAAGTTCTC TCTTATTGTG AGCGAACAGG	1680
TAGTATAACA GAAGCATCAC ACGTTTTCCA AATCTCACGT AATACCATTT ATGGCTGGTT	1740
AAAGCTAAAA GAGAAAACAG GAGAGCTAAA CCACCAAGTA AAAGGAATAA AACCAAGAAA	1800
GGTTGATAGA GATAGACTTA AAAACTATCT TACTGACAAT CCAGACGCTT ATTTGACTGA	1860
AATAGCTTCT GAATTTGGCT GTCATCCAAC TACCATCCAC TATGCGCTCA AAGCTATGGG	1920
LACACTCGAA AAAAAAAGA ACTACACCTA CTATGAAC	1958

## (2) INFORMATION FOR SEQ ID NO: 361:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 851 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

TATGAAATTA AGTTATGATG ATAAAGTTCA GATCTATGAA CTTAGAAAAC AAGGATATAG	60
CTTAGAGAAG CTTTCAAATA AATTGGGAT AmACAATTCT AATCTTAGGT ATATGATTAA	120
ATTGATTGAT CGTTACGGAA TAGAGTTCGT CAAAAAGGA AAAAATCGTT ACTATTCTCC	180
TGATTTAAAA CAAGAAATGA TTAATAAAGT CTGACATGAA GGCTGGACTA AAGATAGAGT	240
TTCTCTTGAA TACGGTCTCC CAAGTCGTAC GATACTTCTT AACTGGCTAG CACAATACAG	300
GAAAAACGGG TATACTATTG TTGAGAAACC AAGAGGGAGA GTACCTGAGA GCGGAGAATG	360
CCATCTTAAA AAAGTTAAGA GAACTCCGAT TGAAGGAGGA AAAAGAGAAA GAAGAAAGAC	420
AGAAATTGTT TAAGAAATTAA TGACTGAGTT TTCGTTAGAT CTTCTTTTAA AAGTCATTAA	480
ACTAGCTCGT TCGACCTACT ACTATCACTT GAAACAGCTA GATAAACCAG ATAAGGACCA	540
AGAGCTTAAA GCTGAAATTC AATCCATTTT TATCGAACAC AAAGGAAATT ATGCTTATCG	600

1375

TCGGATTTAT TTAGAACTAA GAAATCGTGG TTATCTGGTA AATCATAAAA GAGTTCAAGG	660
CTTGATGAAA GTACTCAATc TACAAGCTAA AACGCGACAG AAACGAAAAT ATTCTTCTCA	720
TAAAGGAGAC GTTGGCAAGA AGGCAGAGAA TCTCATTCAA GGCCAATTG AAGGCTCTAA	780
AACAATGGAA CAGTGCTACA CAGATGTGAC AGAATTGGCC ATTCCAGTAA GTACTTAAAA	840
GCTTTACTTA T	851

## (2) INFORMATION FOR SEQ ID NO: 362:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1168 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

GGGTAGAATC GATATCTCCA ATGAGTTGGT CTAGCTGGTG AAAGTGTAAA AAGATTTCGW	60
CCAATTCAAG GTTGAGGCAT CGCAAATAT GGACTGTTTC CTCGTCAGTT CTGGAAAGAA	120
AACGGGATAA GGTGGCTGT GAAGCAAGCT GCCCTCCTTC CAACAATTTT GGAAAGTAGG	180
CATCAGCTGA CAATTCTTTA CAAGCATAGT CCGTTCCATA ACCTGTTAAC AGTTGAAAGA	240
GGAAGTGGAC AAGGATATCT GAATCCGAAT AACGACAGTA GCGGCGTTGG TCATTGCTTA	300
CTAAATACTT AGAAATCCGC TCTTTTAGTT TCAACTGGGA AAAAAGTTCC TGAAAAAGA	360
TAAGACCACC ATACTGGGTT AAATGACCTC CATCGAAAGA TAGTTGGTAA AAAGACTTGT	420
TTTGGAAGTG ATGATTGGT AAAGTGTTC TGTGAGTTTC CTTTCTTTT GTGTTTTTT	480
CTACACTTAT ACCATAAAGG GGAAACTCTT TTTTGTCTAG TAAAAACAC CCATTGGGTG	540
AAAAAAGAAA CCATCCAGGA TCTAAGCTAA GGCAAGGATT CTGGATGGTT TTTAGATTG	600
GGGTGAATAA TTGGGGTTTT AGCTGCTTGC GGCCAATCAG GTTCAGATAC AAAAATTAC	660
TCATCAACCT TTAGTGAAA TCCAATACA TTAACTATC TATTAGACTA TTACGCTGAT	720
AATATAGTCA ATTGAAACAA GAACAAGACA AAAGAGCTC ATAAAAGGTA TTGCAACTTG	780
GTAATACCTT TTTGAGGTGC TTTTGATAT GAGCCCATGT TTTCTCAATA GGATTGTACT	840
CAGGTGAGTA GGGAGGAAGA GGTAAGTT TATACCCAAA CTCTTCACAC AAGAGTTCTA	900
ACTTACCCAT TCTATGGAAT CTTGCATTAT CCATAATAAT AACCGATGGT GTGTTTAATG	960
TTGGTAAGAG AAATTCTGA AACCAAGCTT CAAAAAGTC GCTCGTCATC GTCTCTCGT	1020
AAGTTATTGG AGCGATTAAC TCACATTG TTAGACCTGC AACCAAAGAA ATCCTCTGAT	1080

1376

ATCTTCTTCC AGATACTTTG CCTCTTCTTA ACTGACCTTT TAATGAGCGA CCATATTCTC	1140
GATAAAAATA AGTATCGAAT CCTGTTTC	1168

## (2) INFORMATION FOR SEQ ID NO: 363:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4483 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

GTCAGCTTCA GCAAGCCCAT CAGCTTCTGA ATCTGCATCA ACCAGTGCCT CCGCTTCAGC	60
GTCAACCAGT GCGTCGGCTT CAGCGTCGAC AAGTGCTTCG GCTTCAGCAT CAACGAGTGC	120
GTCGGCCTCA GCAAGCGCAA GTACCTCAGC GTCAGCTTCC GCCTCAACCA GTGCGTCGGC	180
TTCAGCAAGC ACAAGTGCGT CAGCCTCAGC AAGTATCTCA GCGTCTGAAT CGGCATCAAC	240
GAGTGCGTCT GAGTCAGCAT CAACGAGTAC GTCAGCCTCA GCAAGCACAT CAGCTTCTGA	300
ATCTGCATCA ACCAGTGCCT CAGCCTCAGC ATCGACAAGC GCCTCAGCTT CAGCAAGTAC	360
CAGTGCTTCA GCCTCAGCGT CGACAAGTGC GTCGGCCTCA ACCAGTGCAT CTGAATCGGC	420
ATCAACCAGT GCGTCAGCCT CAGCAAGTAC TAGTGCATCA GCTTCAGCAT CAACGAGTGC	480
ATCGGCTTCA GCATCAACCA GTGCCTCGGC TTCAGCGTCA ACCAGTGCCT CAGCTTCAGC	540
AAGTACCAGT GCTTCAGTCT CAGCATCAAC AAGTGCTTCA GCCTCAGCAT CGACAAGTGC	600
CTCGGCTTCA GCAAGCACAT CAGCATCTGA ATCAGCGTCG ACAAGCGCCT CAGCTTCAGC	660
AAGTACCAGT GCGTCAGCCT CAGCGTCGAC AAGTGCGTCA GCCTCAGCAA GTACTAGTGC	720
ATCAGCTTCA GCATCAACGA GTGCATCGGC TTCGGCGTCA ACCAGTGCAT CAGAGTCAGC	780
AAGTACCAGT GCGTCAGCTT CCGCATCAAC AAGTGCTTCG GCTTCAGCAA GCACCAAGTGC	840
GTCGGCTTCA GCAAGTACTA GCGCCTCAGC CTCAGCCTCA ACCAGTGCCT CAGCCTCAGC	900
AAGTATCTCA GCGTCTGAAT CGGCATCAAC GAGTGCGTCC GCTTCAGCAA GTACTAGCGC	960
CTCAGCCTCA GCGTCAACAA GTGCATCGGC TTCAGCGTCA ACGAGTGCCT CTGAATCGGC	1020
ATCAACGAGT GCGTCCGCTT CAGCAAGTAC TAGCGCCTCA GCCTCAGCGT CAACAAGTGC	1080
ATCGGCTTCA GCATCAACGA GTGCGTCCGC TTCAGCAAGT ACTAGCGCCT CAGCCTCAGC	1140
GTCAACAAGT GCATCGGCTT CAGCGTCAAC GAGTGCGTCT GAGTCAGCAT CAACGAGTGC	1200
GTCAGCCTCA GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGCCT CAGCCTCAGC	1260
ATCGACAAGC GCCTCAGCTT CAGCAAGTAC CAGTGCGTCA GCTCAGCGTC GACAAGTGCs	1320



1377

TCrGCTTCAG CAAGTACCAG TGCCTCAGCC TCAGCAAGTA CCAGTGCKTC AGCCTCAGCG	1380
TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG TGCCTCAGCC	1440
TCAGCAAGTA CTAGCGCCTC AGCCTCAGCA TCAACGAGTG CGTCCGCTTC AGCAAGTACT	1500
AGTGTCATCAG CTTAGCAAG TACTAGCGCC TCAGCCTCAG CGTCGACAAG CGCCTCAGCT	1560
TCAGCAAGTA CCAGTGCGTC AGCCTCAGCG TCGACAAGTG CGTCGGCTTC AGCAAGTACC	1620
TCAGCGTCTG AATCAGCATC AACAAAGTGGC TCGGCTTCAG CATCAACGAG TGCATCAGCT	1680
TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG	1740
AGTGCTTCAG TCTCAGCGTC AACCAGTGCC TCTGAATCCG CATCAACAAG TGCCTCGGCT	1800
TCAGCAAGCA CCAGTGCTTC GGCTTCAGCG TCAACGAGTG CGTCTGAGTC AGCATCAACG	1860
AGTGCGTCAC CTCAGCAAGC ACATCAGCTT CTGAATCTGC ATCAACCAGT GCGTCACTTC	1920
CGCATCAACA AGCGCCTCGG CCTCAGCAAG TACAAGTGCT TCAGCCTCAG CATCAACCAG	1980
TGCATCAGCT TCAGCCTCAA CAAGTGCTTC AGCCTCAGCG TCAACGAGTG CCTCGGCTTC	2040
AGCAAGTACC AGTGCGTCAG CTTAGCAAG CACAAGTGCG TCAGCTTCAG CATCAACCAG	2100
TGCTTCGGCT TCGGCATCAA CAAGTGCTTC AGCATCAGCA TCAACGAGTG CGTCACTCA	2160
GCAAGTACTA GTGCATCAGC ATCAGCATCA ACCAGTGCTC CAGCCTCAGC AAGTATCTCA	2220
GCGTCTGAAT CGGCATCAAC GAGTGCTCA GCATCAGCAT CAACGAGTGC ATCGGCTTCA	2280
GCGTCAACCA GTGCATCAGT CTCAGCAAGC ACCAGTGCGT CGGCTTCAGC ATCAACCAGT	2340
GCCTCAGCCT CAGCAAGTAT CTCAGCGTCT GAATCGGCAT CAACGAGTGC GTCAGCTCA	2400
GCAAGTACTA GTGCATCAGC ATCAGCATCA ACCAGTGCTC CGGCTTCAGC AAGTACCAGC	2460
GCCTCAGCTT CAGCAAGCAC CAGTGCGTCA GCCTCAGCAA GTACCAGCGC CTCAGCCTCA	2520
GCAAGCACCA GTGCCTCAGC TTCAGCAAGT ACCAGTGCGT CAGCCTCAGC GTCGACAAGT	2580
GCGTCGGCTT CAGCAAGTAC CTCAGCGTCT GAATCAGCAT CAACGAGTGC ATCAGCTTCA	2640
GCATCAACAA GTGCTTCAGC TTCAGCAAGT ACCAGTGCGT CGGCTTCAGC ATCAACGAGT	2700
GCTTCAGTCT CAGCGTCAAC CAGTGCTCT GAATCAGCAT CAACAAGTGC CTCGGCTTCA	2760
GCAAGCACCA GTGCGTCGGC TTCAGCAAGT ACTAGTGCTC CGGCTTCAGC ATCGACAAGT	2820
GCGTCTGAAT CGGCATCAAC GAGTGCTTCG GCTTCAGCAT CAACGAGTGC GTCAGCCTCA	2880
GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGCGT CCGCTTCAGC GTCAACCAGT	2940
GCGTCGGCTT CAGCGTCGAC AAGTGCTTCG GCTTCAGCAT CAACGAGTGC GTCGGCCTCA	3000
GCAAGCGCAA GTACCTCAGC GTCAGCTTCC GCCTCAACCA GTGCGTCCGC TTCAGCAAGC	3060

1378

ACAAGTGGCT CAGCCTCAGC AAGTATCTCA GCGTCTGAAT CGGCATCAAC GAGTGGCTCG	3120
GCCTCAGCAA GCGCAAGTAC CTCAGCGTCA GCTTCCGCCT CAACCAGTGC GTCGGCTTCA	3180
GCAAGCACAA GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT CTGAATCGGC ATCAACGAGT	3240
GCGTCTGAGT CAGCATCAAC GAGTACGTCA GCCTCAGCAA GCACATCAGC TTCTGAATCG	3300
GCATCAACCA GTGCGTCAGC CTCAGCATCG ACAAGCGCCT CAGCTTCAGC AAGTACCAGT	3360
GCTTCAGCCT CAGCGTCGAC AAGTGGCTCG GCCTCAACCA GTGCATCTGA ATCGGCATCA	3420
ACCAGTGGCT CAGCCTCAGC AAGTACTAGT GCATCAGCTT CAGCATCAAC GAGTGCATCG	3480
GCTTCAGCAT CAACCAGTGC CTCGGCTTCA GCGTCAACCA GTGCGTCAGC TTCAGCAAGT	3540
ACCAGTGCTT CAGTCTCAGC ATCAACAAGT GCTTCAGCCT CAGCATCGAC AAGTGCCTCG	3600
GCTTCAGCAA GCACATCAGC ATCTGAATCA GCGTCGACAA GCGCCTCAGC TTCAGCAAGT	3660
ACCAGTGGCT CAGCCTCAGC GTCGACAAGT GCGTCAGCCT CAGCAAGTAC TAGTGCATCA	3720
GCTTCAGCAT CAACGAGTGC ATCGGCTTCG GCGTCAACCA GTGCATCAGA GTCAGCAAGT	3780
ACCAGTGGCT CAGCTTCCGC ATCAACAAGT GCCTCGGCTT CAGCAAGCAC CAGTGGCTCG	3840
GCTTCAGCAA GTACTAGCGC CTCAGCCTCA GCCTCAACCA GTGCGTCAGC CTCAGCAAGT	3900
ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCCGCTT CAGCAAGTAC TAGCGCCTCA	3960
GCCTCAGCGT CAACAAGTGC ATCGGCTTCA GCGTCAACGA GTGCGTCTGA ATCGGCATCA	4020
ACGAGTGGCT CCGCTTCAGC AAGTACTAGC GCCTCAGCCT CAGCGTCAAC AAGTGCATCG	4080
GCTTCAGCAT CAACGAGTGC GTCCGCTTCA GCAAGTACTA GCGCCTCAGC CTCAGCGTCA	4140
ACAAGTGCAT CCGGTTTCAGC GTCAACGAGT GCGTCTGAGT CAGCATCAAC GAGTGGCTCA	4200
CCTCAKCAAG CACATCAGCT TCTGAATCTG CATCAACCAG TGGCTCACTT CCGCATCAAC	4260
AAGCGCCTCG GCCTCAGCAA GTACAAGTGC TTCAGCCTCA GCATCAACCA GTGCATCAGC	4320
TTCAGCCTCA ACAAGTGCTT CAGCCTCAGC GTCAGACCAG TGCCTCGGCT TCAGCAAGTA	4380
CCAGTGGCTC ACTTCAGCAA GCACAAGTGC GTCAGCTTCA GCATCAACCA GTGCTTCGCT	4440
TTCGGCATCA ACAAGTGCCT CAGCATCAGC ATCAACGAGT GCG	4483

## (2) INFORMATION FOR SEQ ID NO: 364:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2550 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

1379

GTACCTCAGC GTCCTTCCGC CTCAACCAGT GCGTCCGCTT CAGCAAGCAC AAGTGCGTCA	60
CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TCGTCCGCC TCAGCAAGCG	120
CAAGTACCTC AGCGTCACTT CCGCTCAAC CAGTGGCTCG GCTTCAGCAA GCACAAGTGC	180
GTCAsCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGGCTC TGAGTCAGCA	240
TCAACGAGTA CGTCAGCCTC AGCAAGCACA TCAGCTTCTG AATCGGCATC AACCAGTGCG	300
TCAGCCTCAG CATCGACAAG CGCCTCAGCT TCAGCAAGTA CCAGTGCTTC AGCCTCAGCG	360
TCGACAAGTG CGTCGGCCTC AACCAGTGCA TCTGAATCGG CATCAACCAG TCGTCAGCC	420
TCAGCAAGTA CTAGTGATC AGCTTCAGCA TCAACGAGTG CATCGGCTTC AGCATCAACC	480
AGTGCCTCGG CTTCAGCGTC AACCAGTGCG TCAGCTTCAG CAAGTACCAG TGCTTCAGTC	540
TCAGCATCAA CAAGTGCTTC AGCCTCAGCA TCGACAAGTG CCTCGGCTTC AGCAAGCACA	600
TCAGCATCTG AATCAGCGTC GACAAGTGCG TCGGCCTCAA CCAGTGATC TGAATCGGCA	660
TCAACCAGTG CGTCAGCCTC AGCAAGTACT AGTGATCAG CTTCAGCATC AACGAGTGCA	720
TCGGCTTCGG CGTCAACCAG TGCATCAGAG TCAGCAAGTA CCAGTGCGTC AGCTTCGCA	780
TCAACAAGTG CCTCGGCTTC AGCAAGCACA TCAGCATCTG AATCAGCGTC AACCAGTGCT	840
TCGGCTTCAG CAAGTACCAG TGCTTCAGCT TCAGCATCAA CCAGCGCCTC GGCCTCAGCA	900
AGCACCTCAG CTTCTGAATC GGCCTCAACC AGCGCCTCGG CCTCAGCAAG CACCTCAGCT	960
TCTGAATCGG CCTCAACCAG CGCCTCAGCC TCAGCATCAA CGAGTGCTTC GGCTTCAGCA	1020
AGCACAAGCG CCTCGGCTTC AGCATCAACG AGTACGTCAG CTTCAGCGTC AACCAGTGCT	1080
TCAGCCTCAG CATCAACAAG TCGTCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA	1140
TCAACGAGTG CGTCTGAGTC AGCATCAACG AGTACGTCAG CCTCAGCAAG CACAAGTGCT	1200
TCAGCCTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGGCTC CGCTTCAGCA	1260
AGTACTAGCG CCTCAGCATC AGCGTCAACA AGTGCTTCGG CTTCAGCGTC AACGAGTGCG	1320
TCTGAGTCAG CATCAACGAG TACGTCAGCC TCAGCAAGCA CATCAGCTTC TGAATCTGCA	1380
TCAACCAGTG CGTCAGCCTC AGCATCGACA AGCGCCTCAG CTTCAGCAAG TACCAGTGCG	1440
TCAGCCTCAG CAAGTACCAG TGCTTCAGCC TCAGCGTCGA CAAGTGCGTC GGCCTCAACC	1500
AGTGATCTG AATCGGCATC AACCAGTGCG TCAGCTCAGC AAGTACTAGT GCATCAGCTT	1560
CAGCATCAAC GAGTGATCG GCTTCGGCGT CAACCAGTGC ATCAGAGTCA GCAAGTACCA	1620
GTGCGTCAC TCCGCATCAA CAAGTGCTC GGCTTCAGCA AGCACATCAG CATCTGAATC	1680
AGCGTCAACC AGTGCTTCGG CTTCAGCAAG TACCAGTGCT TCAGCTTCAG CATCAACCAG	1740

1380

CGCCTCGGCC TCAGCAAGCA CCTCAGCTTC TGAATCGGCC TCAACCAGCG CCTCGGCCTC	1800
AGCAAGCACC TCAGCTTCTG AATCGGCCTC AACCAGCGCC TCAGCCTCAG CATCAACGAG	1860
TGCTTCGGCT TCAGCAAGCA CAAGCGCCTC GGGTTCAGCA TCAACGAGTA CGTCAGCTTC	1920
AGCGTCAACC AGTGCTTCAG CCTCAGCATC AACAAAGTGG TCAGCCTCAG CAAGTATCTC	1980
AGCGTCTGAA TCGGCATCAA CGAGTGCCTC TGAGTCAGCA TCAACGAGTA CGTCAGCCTC	2040
AGCAAGCACC TCAGCTTCTG AATCGGCCTC AACCAGTGGC TCAGCCTCAG CATCGACAAG	2100
CGCCTCAGCT TCAGCAAGTA CCAGTGCTTC AGCCTCAGCG TCGACAAGTG CGTCGGCCTC	2160
AACCAGTGCA TCTGAATCGG CATCAACCAG TGGTTCAGCC TCAGCAAGTA CTAGTGCATC	2220
GGCTTCAGCA TCAACCAGTG CCTCGGCTTC AGCGTCAACC AGTGCCTCAG CTTTCAGCAAG	2280
TACCACTGCT TCAGTCTCAG CATCAACAAG TGCTTCAGCC TCAGCATCGA CAAGTGCTC	2340
GGCTTCAGCA AGCACATCAG CATCTGAATC AGCGTCGACA AGCGCCTCAG CTTTCAGCAAG	2400
TACCACTGCG TCAGCCTCAG CGTCGACAAG TGGTTCAGCT ACAGCAAGTA CTAGTGCATC	2460
AGCTTCAGCA TCAACGAGTG CATCGGCTTC GGCCTCAACC AGTGCATCAG AGTCAGCAAG	2520
TACCACTGCG TCAGTTCACG CATCAACAAG	2550

## (2) INFORMATION FOR SEQ ID NO: 365:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1436 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 365:

ACCCAGCAAG TACTAGTGCA TCGGCTTCAG CAAGCACCAG TGGTTCGGCT TCAGCATCAA	60
CCAGTGCCTC AGCCTCAGCA AGTATCTCAG CGTCTGAATC GGCATCAACG AGTGCGTCAC	120
CTCAGCAAGT ACTAGTGCAT CAGCATCAGC ATCAACGAGT GCATCGGCTT CAGCAAGTAC	180
CAGCGCCTCA GCTTCAGCAA GCACCAGTGC GTCAsCTCAG CAAGTACCAG CGCCTCAGCC	240
TCAGCAAGCA CCAGTGCCTC AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA	300
AGTGCGTCGG CTTTCAGCAAG TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT	360
TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTATCTCAG CGTCTGAATC GGCATCAACG	420
AGTGCGTCGG CTTTCAGCAAG TACTAGCGCC TCAGCATCAG CGTCAACAAG TGCTTCGGCT	480
TCAGCGTCAA CGAGTGCCTC TGAGTCAGCA TCAACGAGTA CGTCAGCCTC AGCAAGCACA	540
TCAGCTTCTG AATCTGCATC AACCAGTGGC TCAGCCTCAG CATCGACAAG CGCCTCAGCT	600

1381

TCAGCAAGTA CCAGTGCCTC AgCCTCAGCA AGTACCAGTG CTTCAGCCTC AGCGTCGACA	660
AGTGCCTCGG CCTCAACCAG TGCATCTGAA TCGGCATCAA CCAGTGCCTC AGCCTCAGCA	720
AGTACTAGCG CCTCAGCCTC AGCATCAACG AGTGCCTCGG CTTCAGCAAG TACTAGTGCA	780
TCAGCTTCAG CAAGTACTAG CGCCTCAGCC TCAGCGTCGA CAAGCGCCTC AGCTTCAGCA	840
AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCCTCGG CTTCAGCAAG TACCTCAGCG	900
TCTGAATCAG CATCAACAAG TCGCTCGGCT TCAGCATCAA CGAGTGCATC AGCTTCAGCA	960
TCAACAAGTG CTTCAGCTTC AGCAAGTACC AGTGCCTCGG CTTCAGCATC AACGAGTGCT	1020
TCAGTCTCAG CGTCAACCAG TGCTCTGAA TCCGCATCAA CAAGTGCCTC GGCTTCAGCA	1080
AGCACCAGTG CTTGGGCTTC AGCGTCAACG AGTGCCTCGT AGTCAGCATC AACGAGTGCG	1140
TCAGCCTCAG CAAGCACATC AGCTTCTGAA TCTGCATCAA CCAGTGCCTC AGCTTCGCA	1200
TCAACAAGCG CCTCGGCCTC AGCAAGTACA AGTGCTTCAG CCTCAGCATC AACCAGTGCA	1260
TCAGCTTCAG CCTCAACAAG TGCTTCAGCC TCAGCGTCAA CCAGTGCCTC GGCTTCAGCA	1320
AGTACCAGTG CGTCAGCTTC AGCAAGCACA AGTGCCTCAG CTTCAGCATC AACCAGTGCT	1380
TCGGCTTCGG CATCAACAAG TGCTTCAGCA TCAGCATCAA CGAGTGCCTC AGCCGG	1436

(2) INFORMATION FOR SEQ ID NO: 366:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 735 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:

GCAGTTGCCA CACCGTGCTG ACCAGCACCC GTTCTGCGA TAATTTTCTT TTTACCCATG	60
CGTWTGGCAA GCCAACTTG TCCTAAGGCA TTGTTAATCT TGTGGGCTCC TGTATGGTTA	120
AGGTCTTCCC GTTTGAGATA AATCTTGCTC CGCCAATATG CTGGGTCAAG TTTTTTGCGT	180
AATAAAGAGG AGTTTCACGT CCTACGTACT GGCGCAAAAG CTGGTTTAAT TCCTCTTGGA	240
AACTTGGGTC TGCTGACTT TCACGGTAGG CTTCTCCAA CTCCAAACT GCTGTCATCA	300
ATGTTTCTGG GACAAAACGT CCGCCGAATT TTCCGTAAAA TCCATCTTTA TTTGGTTCCT	360
GATATGCCAT GCTTTACCCT CTCTATAAAT CTTCTAATCT TTTCATGATC TTTTGTCCA	420
TCTGTCTCCA CTCCGCTCGA TACATCTACT GCATAGGGAG TAAAGTGTTG AATTGCTTTT	480
ACTACATTAT CTTCAATTAAG GCCACCTGCG ATAAAGAAGG GCTGTGCTAG TCCAGTCGTA	540

1382

TCCAGTTGAC	CCCAATCAAA	GGGCTGGCCA	CTTCCTGCCA	CAGGGGCATC	AAAGAGTAGA	600
TAATCTGCCT	GAGAATTGGG	GACATGCCCA	TTTCATCTA	CCTGCACAGC	CTGAATACTG	660
GCACAAGGCA	AATTCTCAAA	TAAATCATCT	GCCACCTGAC	CGTGAACCTG	AACCAAGTCC	720
AAGCCGGGGA	TCCTC					735

## (2) INFORMATION FOR SEQ ID NO: 367:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1702 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:

TACTAGCGCC	TCAGCCTCAG	CGTCAACAAG	TGCATCGGCT	TCAGCATCAA	CGAGTGCCTC	60
CGCTTCAGCA	AGTACTAGCG	CCTCAGCCTC	AGCGTCAACA	AGTGCATCGG	CTTCAGCGTC	120
AACGAGTGGC	TCTGAGTCAG	CATCAACGAG	TGCGTCAGCC	TCAGCAAGCA	CATCAGCTTC	180
TGAATCTGCA	TCAACCAGTG	CGTCAGCCTC	AGCATCGACA	AGCGCCTCAG	CTTCAGCAAG	240
TACCAGTGGC	TCAGCCTCAG	CGTCGACAAG	TGCGTCGGCT	TCAGCAAGTA	CCAGTGCCTC	300
AGCCTCAGCA	AGTACCAGTG	CGTCAGCCTC	AGCGTCGACA	AGTGCCTCGG	CCTCAACCAG	360
TGCATCTGAA	TCGGCATCAA	CCAGTGCCTC	AGCCTCAGCA	AGTACTAGTG	CATCAGCTTC	420
AGCATCAACG	AGTGCATCGG	CTTCAGCATC	AACCAGTGCA	TCAGAGTCAG	CAAGTACCAG	480
TGCGTCAGCT	TCGGCATCAA	CAAGTGCTCT	GGCTTCAGCA	AGTACTAGCG	CCTCAGCCTC	540
AGCGTCAACA	AGTGCTTCAG	CTTCGCGCTC	AACCAGCGCC	TCGGCCTCAG	CAAGTATCTC	600
AGCGTCTGAA	TCGGCATCAA	CAAGTGCTCT	GGCTTCAGCA	TCAACGAGTG	CATCAGTCTC	660
AGCAAGCACC	AGTGCGTCGG	CCTCAGCAAG	CACCAGCGCG	TCTGAATCCG	CATCAACCAG	720
TGCCTCAGCT	TCAGCAAGTA	CCTCAGCATC	TGAATCAGCA	TCAACAAGTG	CATCGGCTTC	780
AGCAAGCACA	AGTGCTTCAG	CCTCAGCAAG	TATCTCAGCG	TCTGAATCCG	CATCAACGAG	840
TGCGTCCGCT	TCAGCAAGTA	CTAGCGCCTC	AGCATCAGCG	TCAACAAGTG	CTTCGGCTTC	900
AGCGTCAACG	AGTGCGTCTG	AGTCAGCATC	AACGAGTACG	TCAGCCTCAG	CAAGCACATC	960
AGCTTCTGAA	TCTGCATCAA	CCAGTGCCTC	AGCCTCAGCA	TCGACAAGCG	CCTCAGCTTC	1020
AGCAAGTACC	AGTGCGTCAG	CCTCAGCAAG	TACCAGTGCT	TCAGCCTCAG	CGTCGACAAG	1080
TGCGTCCGCC	TCAACCAGTG	CATCTGAATC	GGCATCAACC	AGTGCGTCAG	CCTCAGCAAG	1140
TACTAGCGCC	TCAGCCTCAG	CATCAACGAG	TGCGTCCGCT	TCAGCAAGTA	CTAGTGCATC	1200

1383

AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC AGCGCCTCAG CTTCAGCAAG	1260
CACCACTGCG TCAGCCTCAG CAAGTACCAG CGCCTCAGCC TCAGCAAGCA CCAGTGCCTC	1320
AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA AGTGCGTCGG CTTCAGCAAG	1380
TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT TCAGCATCAA CAAGTGCTTC	1440
AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG AGTGCTTCAG TCTCAGCGTC	1500
AACCACTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT TCAGCAAGCA CCAGTGCCTC	1560
GGCTTCAGCA AGTACTAGTG CATCGGCTTC AGCATCGACA AGTGCGTCTG AATCGGCATC	1620
AACGAGTGCT TCGGCTTCAG CATCAACGAG TGCCTCAGCC TCAGCAAGCA CATCAGCTTC	1680
TGAATCTGCA TCAACCACTG CG	1702

(2) INFORMATION FOR SEQ ID NO: 368:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 941 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

ACCACTGCAT CAGCTTCAGC CTCAACAAGT GCTTCAGCCT CAGCGTCAAC CAGTGCCTCG	60
GCTTCAGCAA GTACCACTGC GTCACCTCAG CAAGCACAAG TCGGTCACTT CAGCATCAAC	120
CAGTGCTTCG GCTTCGGCAT CAACAAGTGC CTCAGCATCA GCATCAACGA GTGCGTCACC	180
TCAGCAAGTA CTAGTGCATC AGCATCAGCA TCAACCACTG CATCAGCCTC AGCAAGTATC	240
TCAGCGTCTG AATCGGCATC AACGAGTGCA TCAGCATCAG CATCAACGAG TGCATCGGCT	300
TCAGCGTCAA CCAGTGCATC AGTCTCAGCA AGCACCAGTG CGTCGGCTTC AGCATCAACG	360
AGTGCTTCAG CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TGCCTCAGCC	420
TCAGCAAGTA CTAGTGCATC GGCTTCAGCA AGCACCAGTG CGTCGGCTTC AGCATCAACC	480
AGTGCTTCAG CCTCAGCAAG TATCTCAGCG TCTGAATCGG CATCAACGAG TGCCTCAGCC	540
TCAGCAAGTA CTAGTGCATC AGCATCAGCA TCAACGAGTG CATCGGCTTC AGCAAGTACC	600
AGCGCCTCAG CTTCAGCAAG CACCACTGCG TCAGCCTCAG CAAGTACCAG CGCCTCAGCC	660
TCAGCAAGCA CCAGTGCCTC AGCTTCAGCA AGTACCAGTG CGTCAGCCTC AGCGTCGACA	720
AGTGCGTCGG CTTCAGCAAG TACCTCAGCG TCTGAATCAG CATCAACGAG TGCATCAGCT	780
TCAGCATCAA CAAGTGCTTC AGCTTCAGCA AGTACCAGTG CGTCGGCTTC AGCATCAACG	840

1384  
AGTGCTTCAG TCTCAGCGTC AACCAGTGCC TCTGAATCAG CATCAACAAG TGCCTCGGCT 900  
TCAGCAAGCA CCAGTGCCTC GGCTTCAGCA AGTACTAGTG C 941

(2) INFORMATION FOR SEQ ID NO: 369:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 869 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

CAGCAAGTAC TAGTGCATCA GCTTCAGCAT CAACGAGTGC ATCGGCTTCT GCGTCAACCA 60  
GTGCATCAGA GTCAGCAAGT ACCAGTGCCT CAGCTTCCGC ATCAACAAGT GCCTCGGCTT 120  
CAGCAAGCAC CAGTGCCTCG GCTTCAGCAA GTACTAGCGC CTCAGCCTCA GCCTCAACCA 180  
GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCCGCTT 240  
CAGCAAGTAC TAGCGCCTCA GCCTCAGCGT CAACAAGTGC ATCGGCTTCA GCGTCAACGA 300  
GTGCGTCTGA ATCGGCATCA ACGAGTGCCT CCGCTTCAGC AAGTACTAGC GCCTCAGCCT 360  
CAGCGTCAAC AAGTGCATCG GCTTCAGCAT CAACGAGTGC GTCCGCTTCA GCAAGTACTA 420  
GCGCCTCAGC CTCAGCGTCA ACAAGTGCAT CGGCTTCAGC GTCAACGAGT GCGTCTGAGT 480  
CAGCATCAAC GAGTGCCTCA GCCTCAGCAA GCACATCAGC TTCTGAATCT GCATCAACCA 540  
GTGCGTCAGC CTCAGCATCG ACAAGCGCCT CAGCTTCAGC AAGTACCACT GCGTCAGCCT 600  
CAGCGTCGAC AAGTGCCTCG GCTTCAGCAA GTACCAAGTGC GTCAGCCTCA GCAAGTACCA 660  
GTGCGTCAGC CTCAGCGTCT ACAAGTGCCT CGGCCTCAAC CAGTGCATCT GAATCGGCAT 720  
CAACCAAGTGC GTCAGCCTCA GCAAGTACTA GTGCATCAGC TTCAGCATCA ACGAGTGCAT 780  
CGGCTTCAGC ATCAACCACT GCATCAGAGT CAGCAAGTAC CAGTGCCTCA GNTTCCGCAT 840  
GCAACAAGTG CTTCCGCTTC AGCAAGTAC 869

(2) INFORMATION FOR SEQ ID NO: 370:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 750 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

TCAACAAGTG CCTCAGCATC AGCATCAACG AGTGCCTCAG CCTCAGCAAG TACTAGTGCA 60



1385

TCAGCATCAG CATCAACCAG TGCATCAGCC TCAGCAAGTA TCTCAGCGTC TGAATCGGCA	120
TCAACGAGTG CATCAGCATC AGCATCAACG AGTGCATCGG CTTCAGCGTC AACCAGTGCA	180
TCAGTCTCAG CAAGCACCAG TGCCTCGGCT TCAGCATCAA CGAGTGCCTC AGCCTCAGCA	240
AGTATCTCAG CGTCTGAATC GGCATCAACG AGTGCCTCAG CCTCAGCAAG TACTAGTGCA	300
TCGGCTTCAG CAAGCACCAG TGCCTCGGCT TCAGCATCAA CCAGTGCCTC AGCCTCAGCA	360
AGTATCTCAG-CGTCTGAATC-GGCATCAACG-AGTGCCTCAG-GCTCAGCAAG-TACTAGTGCA-	420
TCAGCATCAG CATCAACGAG TGCATCGGCT TCAGCAAGTA CCAGCGCCTC AGCTTCAGCA	480
AGCACCAGTG CGTCAGCCTC AGCAAGTACC AGCGCCTCAG CCTCAGCAAG CACCAGTGCC	540
TCAGCTTCAG CAAGTACCAG TGCCTCAGCC TCAGCGCTCGA CAAGTGCCTC GGCTTCAGCA	600
AGTACCTCAG CGTCTGAATC AGCATCAACG AGTGCATCAG CTTCAGCATC AACAAGTGCT	660
TCAGCTTCAG CAAGTATCTC AGCGTCTGAA TCGGCATCAA CGAGTGCCTC CGCTTCAGCA	720
AGTACTAGCG CCTCAGCATC AGCGTCAACG	750

(2) INFORMATION FOR SEQ ID NO: 371:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 957 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

CCGGAAAACA GCTCTGGCGC TTGGTCTTGC CCAGCGTATT GCTAGTGGTG ACGTGCCTGC	60
GGAAATGGCT AAGATGCCCG TGTTAGAACT TGATTGTATG AATGTCGTTG CAGGGACACG	120
CTTCCGTGGT GACTTTGAAG AACGCATGAA TAATATCATC AAGGATATTG AAGAAGATGG	180
CCAAGTCATC CTCTTTATCG ATGAACTCCA CACCATCATG GGTTCCTGGTA GCGGGATTGA	240
TTGACTCTG GATGCGGCCA ATATCTTGAA ACCAGCCTTG GCGCGTGGAA CTTTGAGAAC	300
GGTTGGTGCC ACTACTCAGG AAGAATATCA AAAACATATC GAAAAAGATG CGGCACTTTC	360
TCGTCGTTTC GCTAAAGTGA CGATTGAAGA ACCAAGTGTG GCAGATAGTA TGAATATTTT	420
ACAAGGTTTG AAGGCGACTT ATGAGAAACA TCACCGTGTA CAAATCACAG ATGAAGCGGT	480
TGAAACAGCG GTTAAGATGG CTCATCGTTA TTAAACAGT CGTCACTTGC CAGACTCTGC	540
TATCGATCTC TTGGATGAGG CGGCAGCAAC AGTGCAAAAT AAGGCAAAGC ATGTAAAAGC	600
AGACGATTCA GATTGAGTC CAGCTGACAA GGCCCTGATG GATGGCAAGT GGAAACAGGC	660

1386

AGCCCAGCTA ATCGCAAAAG AAGAGGAAGT ACCTGTCTAC AAAGACTTGG TGACAGAGTC	720
TGATATTTTG ACCACCTTGA GTCGCTTGTC AGGAATCCCA GTTCAAAAAC TGA CTCAAAC	780
GGATGCTAAG AAGTATTTAA ATCTTGAAGC AGAACTCCAT AAACGGGTTA TCGGTCAAGA	840
TCAAGCTGTT TCAAGCATTG GCCGTGCCAT TCGCCGCAAC CAGTCAGGGA TTCGAGTCA	900
TAAGCGTCCG ATTGGTTCCT TTATGTTCTT AGGGCCTACA GGTGTCGGGG TATCCGA	957

## (2) INFORMATION FOR SEQ ID NO: 372:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 807 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

CAAAGCGCCT CAGCTTCAGC ATCAACAAGT GCGTCGGCTT CAGCATCAAC CAGTGCCTCG	60
GCTTCAGCGT CAACCAGTGC GTCACATTCA GCAAGTACCA GTGCTTCAGT CTCAGCATCA	120
ACAAGTGCTT CAGCCTCAGC ATCGACAAGT GCCTCGGCTT CAGCAAGCAC ATCAGCATCT	180
GAATCAGCGT CAACCAGTGC TTCGGCTTCA GCAAGTACCA GTGCTTCAGC TTCAGCATCA	240
ACCAGCGCCT CGGCCTCAGC AAGCACCTCA GCTTCTGAAT CGGCCTCAAC CAGCGCCTCG	300
GCCTCAGCAA GCACCTCAGC TTCTGAATCG GCCTCAACCA GCGCCTCAGC CTCAGCATCA	360
ACGAGTGCTT CGGCTTCAGC AAGCACAAGC GCCTCGGGTT CAGCATCAAC GAGTACGTCA	420
GCTTCAGCGT CAACCAGTGC TTCAGCCTCA GCATCAACAA GTGCGTCAGC CTCAGCAAGT	480
ATCTCAGCGT CTGAATCGGC ATCAACGAGT GCGTCTGAGT CAGCATCAAC GAGTACGTCA	540
GCCTCAGCAA GCACCTCAGC TTCTGAATCG GCCTCAACCA GTGCGTCAGC CTCAGCATCG	600
ACAAGCGCCT CAGCTTCAGC AAGTACCAGT GCTTCAGCCT CAGCGTCGAC AAGTGCCTCG	660
GCCTCAACCA GTGCATCTGA ATCGGCATCA ACCAGTGGT CAGCCTCAGC AAGTACTAGT	720
GCATCGGCTT CAGCATCAAC CAGTGCCTCG GCTTCAGCGT CAACCAGTGC GTCAGCTTCA	780
GCAAGTACCA TGTGCTTCAT GTCTCAG	807

## (2) INFORMATION FOR SEQ ID NO: 373:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1068 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

1387

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

CATCGGCTTC AGCATCAACG AGTGGCTCCG CTTCAGCAAG TACTACCGCC TCAGCCTCAG	60
CGTCAACAAG TGCATCGGCT TCAGCGTCAA CGAGTGGCTC TGAGTCAGCA TCAACGAGTG	120
CGTCACCTCA GCAAGCACAT CAGCTTCTGA ATCTGCATCA ACCAGTGGCT CACCTCAGCA	180
TCGACAAGCG CCTCAGCTTC AGCAAGTACC AGTGGCTCAC CTCAGCGTCG ACAAGTGGCT	240
CGGCTTCAGC AAGTACCAGT GCGTCAsCTC AGCAAGTACC AGTGGCTCAC CTCAGCGTCG	300
ACAAGTGGCT CGGCCTCAAC CAGTGCATCT GAATCGGCAT CAACCAAGTGC GTCACCTCAG	360
CAAGTACTAG TGCATCAGCT TCAGCATCAA CGAGTGCATC GGCTTCAGCA TCAACCAAGTG	420
CATCAGAGTC AGCAAGTACC AGTGGCTCAG CTTCGGCATC AACAAGTGCC TCGGCTTCAG	480
CAAGTACTAG CGCCTCAGCC TCAGCGTCAA CAAGTGCTTC AGCTTCCGGC TCAACCAAGCG	540
CCTCGGCCTC AGCAAGTATC TCAGCGTCTG AATCGGCATC AACAAGTGCC TCGGCTTCAG	600
CATCAACGAG TGCATCAGTC TCAGCAAGCA CCAGTGGCTC GGCCTCAGCA AGCACCAGCG	660
CGTCTGAATC CGCATCAACC AGTGCCTCAG CTTCAGCAAG TACCTCAGCA TCTGAATCAG	720
CATCAACAAG TGCATCGGCT TCAGCAAGCA CAAGTGCTTC AGCCTCAGCA AGTATCTCAG	780
CGTCTGAATC GGCATCAACG AGTGGCTCCG CTTCAGCAAG TACTAGCGCC TCAGCATCAG	840
CGTCAACAAG TGCTTCGGCT TCAGCGTCAA CGAGTGGCTC TGAGTCAGCA TCAACGAGTA	900
CGTCAGCCTC AGCAAGCACA TCAGCTTCTG AATCTGCATC AACCAGTGGC TCAGCCTCAG	960
CATCGACAAG CGCCTCAGCT TCAGCAAGTA CCAGTGGCTC AGCCTCAGCA AGTACCAGTG	1020
CTTCAGCCTC AGCGTCGACA AGTGGCTCGG GCTCAACCAG TGCATCTG	1068

## (2) INFORMATION FOR SEQ ID NO: 374:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 620 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

CAGCATCAAC GAGTGCTTCA GTTTCAGCGT CAACCAAGTG CTCTGAATCA GCTTCAACAA	60
GTGCCTCGGC TTCAGCAAGC CCCAGTGGCT CGGCTTCAGC AAGTACTAGT GCATCGGCTT	120
CAGCATCGAC AAGTGGCTCT GAATCGGCAT CAACGAGTGC TTCGGCTTCA GCATCAACGA	180
GTGCGTCAGC CTCAGCAAGC ACATCAGCTT CTGAATCTGC ATCAACCAGT GCGTCCGyTT	240

1388

CAGCGTCAAC CAGTGCCTCG GCTTCAGCGT CGACAAGTGC TTCGGCTTCA GCATCAACGA	300
GTGCGTCGGC CTCAGCAAGC GCAAGTACCT CAGCGTCAGC TTCCGCCTCA ACCAGTGCCT	360
CGGCTTCAGC AAGCACAAGT GCGTCAGCCT CAGCAAGTAT CTCAGCGTCT GAATCGGCAT	420
CAACGAGTGC GTCTGAGTCA GCATCAACGA GTACGTCAGC CTCAGCAAGC ACATCAGCTT	480
CTGAATCTGC ATCAACCAGT GCGTCAGCCT CAGCATCGAC AAGCGCCTCA GCTTCAGCAA	540
GTACCAGTGC TTCAGCCTCA GCGTCGACAA GTGCGTCGGC CTCAACCAGT GCATCTGAAT	600
CGGCATCAAC CAGTGCCTCA	620

(2) INFORMATION FOR SEQ ID NO: 375:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 720 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

GTATTGGGGC GCCCAACCT CTATGTGACT ACGGATTATT TCCTAGATTA CATGgGGATA	60
AACCATTAG AAGAATTACC AGTGATTGAT GAGCTTGAGA TTCAAGCCCA AGAAAGCCAA	120
TTATTTGGTG AAAGGATAGA AGAAGATGAG AATCAATAAG TATATTGCCC ACGCAGGTGT	180
GGCCAGTAGG AGAAAAGCAG AAGAGCTGAT TAAGCAAGGC TTGGTGACGG TTAACGGCCA	240
AGTGGTGGCT GAACTAGCAA CCACTATCAA GTCAGGCGAC AAGTCGAAG TTGAAGTCA	300
ACCTATCTAC AACGAAGAAA AGGTCTACTA TCTGCTTAAC AAACCACGGG GTGTGATTTC	360
CAGTGTGACA GATGATAAGG GTCGCAAGAC GGTGTGCGAC CTCTTGCCCA ATGTCAAAGA	420
GCGTATTAC CCTGTGGGTC GTTTGGACTG GGATACATCA GGTGTCTTGA TTTTGACCAA	480
TGATGGGGAC TTTACAGACG AGATGATTCA CCCTCGTAAT GAGATTGACA AGGTTTATGT	540
CGCGCGTGTT AAAGGTGTGG CCAATAAGGA CAATCTCCGC CCCTTGACCC GTGGTCTTGA	600
GATTGATGGT AAGAAAACCA AGCCATAATA TATAGGTTTT GTAGCCTCTA CACCATAAAT	660
ATTTGCTAAT AAAAATACTG TATTATTACC CTCTTAAGGT GCGAAATTAT TCAAGTTCTT	720

(2) INFORMATION FOR SEQ ID NO: 376:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 648 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

1389

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

CGCCATTTCC CATCGTACCG CCGAAAATCC CAGCGCCTCA GCCATCAAAT ATCCTATCAA	60
CGTTCTCAAA AAAAGTGACC GCTCTCTCAT CATGTTTCCA AGTGGTAGCC GCCACTCAAA	120
CGATGTCAAG GGGGGCGCAC ACTskATTGC CAAAATGGCC AAGGTCCGTA TCATGCCCGT	180
TACCTACACC GGTCCCATGA CTTTGAAGGG CTTGATTAGC CGTGAACGTG TCGATATGAA	240
CTTTGGAAAT CCAATCGATA TCTCAGATAT CAAGAAAATG AATGATGAAG GCATTGAAAC	300
AGTCGCCAAT CGTATTCAAA CAGAATTCCA ACGTCTGGAC GAAGAAACGA AACAATGGCA	360
CAATGATAAA AAACCAAATC CACTCTGGTG GTTTATCCGC ATCCCTGCCC TCATCCTTGC	420
TATTATCCTC GCTATCCTAA CCATCATCTT TAGCTTTATC GCAAGCTTCA TCTGGAACCC	480
AGATAAGAAA AGAGAAGAAC TTGCATAGAA GAAATGAACC TTGGCCAAAC AGCTAAGGTT	540
TTCATTTATA TAGTAGATTG GwACTAGAAT AGTACACCTC TACTTCTAAA ACATTTTATG	600
AAATCGATTT GACTGTCCTG ATCGATTGT CCTAATCTTA TTTCAATT	648

## (2) INFORMATION FOR SEQ ID NO: 377:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 690 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

GTGCATCGCT TTCAGCATCG ACAAGTGGCT CTGAATCGGC ATCAACGAGT GCTTCGGCTT	60
CAGCATCAAC GAGTGGCTCA GCTTCAGCAA GCACATCAGC TTCTGAATCT GCATCAACCA	120
GTGCGTCCGC TTCAGCGTCA ACCAGTGGCT CGGCTTCAGC GTCGACAAGT GCTTCGGCTT	180
CAGCATCAAC GAGTGGCTCG GCCTCAGCAA GCGCAAGTAC CTCAGCGTCA GCTTCCGCCT	240
CAACCAGTGC GTCCGCTTCA GCAAGCACAA GTGCGTCAGC CTCAGCAAGT ATCTCAGCGT	300
CTGAATCGGC ATCAACGAGT GCGTCGGCCT CAGCAAGCGC AAGTACCTCA GCGTCAGCTT	360
CCGCCTCAAC CAGTGGCTCG GCTTCAGCAA GCACAAGTGC GTCAGCCTCA GCAAGTATCT	420
CAGCGTCTGA ATCGGCATCA ACGAGTGGCT CTGAGTCAGC ATCAACGAGT ACGTCAGCCT	480
CAGCAAGCAC ATCAGCTTCT GAATCGGCAT CAACCAGTGC GTCAGCCTCA GCATCGACAA	540
GCGCCTCAGC TTCAGCAAGT ACCAGTGCTT CAGCCTCAGC GTCGACAAGT GCGTCGGCCT	600
CAACCAGTGC ATCTGAATCG GCATCAACCA GTGCGTCAGC CTCAGCAAGT ACTAGTCAT	660

CAGCTTCAGC ATCAACGAGT GCATCGGCTT 1390 690

(2) INFORMATION FOR SEQ ID NO: 378:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

CGAGATTCTC TGGAGTTATG GATGTCGTTC CAATATGTGC ACGTTGGAAT GTTAGTGCTT 60  
ATATGGGGGG AACAGAATCC TCTCTTGATT GAAGACAAGC TAGTCATTAG GCTGGTTTGT 120  
CTTTTGTCA ACTGTAGTGG GTTGATATAA TAGTATTAGT GAGTGGGATA AAAGTTTCAT 180  
TTAGTTTATT CAGTACAAAT TTAACGGGTC AAGATTTATA TACTAGTGGT GTTTTTGGGG 240  
CTGAGAGAAG TATCTTGATT TTATGTGTGG TTTTATACT TACAGTTGTT CTGCTCCAAA 300  
GAGCTTGTAG AGAAGAATTA GCTCATAAAG GAGATTGATT ATTTTGATAT CAAAAAATG 360  
CACAGGATAA CCTGATGCAT TTTTITAGCG ACAATGCTTG CTACTTCCTT CTGTCGAATT 420  
TAGACAATTT TAAACCCCAA TTATTCACCC CAAATCTAAA AACCATCCAG AATCCTTGCC 480  
TTAGCTTAGA TCCTGGATGG TTTCTTTTTT CACCCAATGG GTGTTTTTTA CTAGACAAAA 540  
AAGAGTTTCC CCTTTATGGT ATAAGTGTAG AAAAAAACAC AAAAAGAAAG GAAACTCACA 600  
TGAACAGTTT ACCAAATCAT CACTTCCAAA ACAAGTCTTT TTACCAACTA TCTTTCGATG 660  
GAGGTCATTT AACCCAGTAT GGTGGTCTTA TCTTTTTTCA GGAACTTTTT TCCCAGTTGA 720  
AACTAAAAGA GCGGATTCTT AAGTATTTAG TAACGAATGA CCAACGCCGC TACTGTCGTT 780  
ATTCGGATTC AGATATCCTT GTCCAGTTCC TCTTCAACT GTTAACAGGT TATGGAACGG 840  
ACTATGCTTG TAAAGAATTG TCAGCTGATG CCTACTTTCC AAAATTATTG GAAGGAGGGC 900  
AGCTTGCTTC ACAGCCAACC TTATCCCGTT TTCTTCCAG AACTGACGAG GAAACAGTCC 960  
ATAGTTTGCG ATGCCTCAAC CTTGAATTGG TCGAATTCTT TTT 1003

(2) INFORMATION FOR SEQ ID NO: 379:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 738 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

1391

CCGATGATTC TGATTGGTTT GCTCTTTACT TTGCTGGGAA TTTTGAGGTA GATCTATGAT	60
TGAAATACTA ATTGTTTTAG CTATTATCCT ATCTCTTGCT TTGATTGTAT TGGTAACTAT	120
ACAACCCCGT CAAAATCAAC TATTTTCCAT GGATGCCACT AGTAATATTG GTAAACCAAG	180
CTACTGGCAG AGCAACACCT TGGTCAAGGT GCTCACTTTA TTGGTGAGTT TGGCTTTATT	240
TATTCTACTA TTAACCTTTA TGGTGATTAC TTATAAATAA AAGAAAACCT CAGATATTCA	300
<u>CCTTTTGTGG ATTGGTCTGA AGTTTTCTTT TTTATACTCA ATGAAAATCA AAGACCAAAC</u>	360
TAGGAAGCTA GCCGCAcKGC TCAAAACACC GTTTTGAGGT TGTAATATA ACTGACGAGc	420
GACTCAAAAC ACCGTTTTGA GGTGTAGAT ATAAGTACG AGcGACTCAA AACACCGTTT	480
TGAGGTTGTG GATAGAACTG ACGAGcGACT CAAAACACCG TTTTGAGGTT GTGGATAGAA	540
CTGACGAAGT CGcTCAAAAC ACCGTTTTGA GGTGTGGAT AGAACTGACG AatgctCAAA	600
ACACCGTTTT GAGGTTGTGG ATAGAACTGA CGAAGCgaac ATATATACAG CAAGGCGACG	660
CTGACGTGGT TTGAAGAGTA TTACTGTCTA TATTTTTGGT AAAAATCAAC TTTTACTTGG	720
ATGAAGGTTT TTTTTTTT	738

(2) INFORMATION FOR SEQ ID NO: 380:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 695 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

CCGTCTTATC AAAGAGGTTA ACAAAGGCAC CAAATTTCTC GATACGAACG ACTTTAGCAC	60
GGTAAACTTC ATCCACTTTG GCTTCACGAA CCAAACCAGC AATAATTTCT TTGGCACGGT	120
TAATAGCATC TTGGTCACTA GAGTAGATAG ACACATTTC TTCTTCGTCT ATATCAATCT	180
TAAACCTGT TTCAGCGATA ATCTTGTCGA TGGTTTCTCC ACCCTTACCG ATGACAATCT	240
TAATCTTGTC CACATCAATC TTGATCGTAT CAATTTTCGG AGCAGTTGGA GCCAATTCTG	300
GACGAACCTC TGAATGGTT GCTTCAATGA CATCAAGGAT TTCAAAACGC GCTTTCTTGG	360
CTTGAGCAAG AGCCTCCGTC AAGATTTCG CAGTAATCCC TTGAATCTTG ATATCCATTT	420
GAAGGGCTGT AATCCCATCA CGAGTACCTG CAACCTTGAA GTCCATATCT CCAAAGTGAT	480
CTTCCAAACC TTGGATATCT GTCAATACTG TGTAGTTATT TCCATCTGAG ATAAGCCCCA	540
TAGCAATACC AGCTACTGGC GCCTTGATTG GCACACCACC AGCCATAAGG GCAAGAGTTC	600

1392  
CCGCACAGAT AGAAGCTTGA GATGAAGAAC CGTTTGATTC CAAAACCTTCT GCTACTAGAC 660  
GGATAGCGTA GGGGAATTCT TCCAAGCTTG GCAGG 695

(2) INFORMATION FOR SEQ ID NO: 381:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 691 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

GACATCTTAT CTAAATACAT GCTAATATAT TTAGATACAA ACATTCCAAC TTGATAATT 60  
TCACTCATCT TTCATCATTC CTTATACAAC TATGCAGTAT AAATAGAATA GTTTTCTCAT 120  
CAGAATGAGA CTATTTTAAT ATTAGATCCC CAATTATPCA CCCCAAATCT AAAAACCATC 180  
CAGAATCCTT GCCTTAGCTT AGATCCTGGA TGGTTTCTTT TTTCACCCAA TGGGTGTTTT 240  
TTACTAGACA AAAAAGAGTT TCCCCTTTAT GGTATAAGTG TAGAAAAAAA CACAAAAAGA 300  
AAGGAACTC ACATGAACAG TTTACCAAT CATCACTTCC AAAACAAGTC TTTTACCAC 360  
CTATCTTTCG ATGGAGGTCA TTTAACCAG TATGGTGGTC TTATCTTTTT TCAGGAACCT 420  
TTTTCCCAGT TGAAGCTAAA AGAGCGGATT TCTAAGTATT TAGTAACGAA TGACCAACGC 480  
CGCTACTGTC GTTATTCGGA TTCAGATATC CTTGTCCAGT TCCTCTTTCA ACTGTTAACA 540  
GGTTATGGAA CGGACTATGC TTGTAAAGAA TTGTCAGCTG ATGCCTACTT TCCAAAATTG 600  
TTGGAAGGAG GGCAGCTTGc TTCACAGCCA ACCTTATCCC GwTTTCTTTC CAGAACTGAC 660  
GAGGAAACAG TCCATAGTTT GCGATGCCTC A 691

(2) INFORMATION FOR SEQ ID NO: 382:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 750 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

ATCTCTCTGC GTAATGGTCC TCAGATAACT CTGATGATGT GTGGCGATAT AGAACTGAGC 60  
CAAGTTATGC CTAAAGGGCC TTAGGAATAG GAGCTTTCAC AAGCTTATCC AGATGATTAT 120  
CTTTTACTCG TTATGGACAA TGCTATATGG CATAAATCAA GTACCTTAAA GATTCCGACT 180  
AATATTGGCT TTGCATTTAT TCCTCCATAC ACACCAGAGA TGAACCCCAT TGAACAAGTG 240



1393

TGGAAAGAGA TTCGTAAACG TGGATTTAAG AATAAAGCCT TTCGAACCTT GGAAGATGTC	300
ATACAAGGAC TGGAGAAGGA GGTGATAAAG TCCATCGTTA ATCGGAGACG GACTAGAATG	360
CTTTTTGAAA ACAGATGAGT ATAAAAAGAA AGTCCTCATT TCAATAGAAA TCACGACTTT	420
CTGATGAATT TATAGTAAAA TGAAATAAGA ACAGGATAGT CAAATCGATT TCTAACAATG	480
TTTTAGAAGC AGAGGTGTAC TATTCTAGTT TAAATCCACT ATATTTGGGG AGTGATAGAA	540
AAGCCCTTCA TCAGCCAATC TACTTGTTCA GGTGCGAGAG CTTTGACATC CTTTCTGTGA	600
CTGGACCAAG TCAGTTTTCC GTTCTCAAAG CGTTTATATA ATATCCAAAA TCCTTGACCA	660
TCCCAGTAAA GAACTTTAAA GCGGTCTTTA CGTCCACCAC AAAAGAGAAA GACTTGATCG	720
GAGAAAGGAT CCAATTCAAA GTGGGTTTGG	750

(2) INFORMATION FOR SEQ ID NO: 383:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 738 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

TCAAATTCCTT CGTGGTCCGC ATATCTnTCT TCGTACACGG CAGTCACTTG GTCTTTCAC	60
ACTCGAGTCG CAGCTTCACG GGCCAATTC TCTTCTACTT GAACTGCCTT TTGGAGGTCA	120
CTGTTGTAGG CTGCAATGAT TTCAGCTTGC AATTCAGCAT CCACGTGAAG CAATTCCACT	180
TCTGCTTTT CTTTACCGAC AGCAGCAACG ATTTCTTCTT GGAAGGCAAT CAATTCTTTG	240
ACAGCTTCGT GCCCTTTAAG GAGCGCTTCC AACATGATTT CTTCTGACAA TTCTTTGGCA	300
CCAGACTCTA CCATGTTGAT AGCGTGCTTG GTTCCAGCTA CTGTCAATTC AAGAAGAGAT	360
TGCTCTGCTT GTTCTTGACT TGGGTTGATG ATGATTTGGC CATCTACATA TCCCACCTGT	420
ACCCAGCAA TTGGTCCGTC AAATGGAATA TCTGAAATAG ACAGTGCCAA AGATGAACCA	480
AACATAGCAG CCATTGGTGC AGATGCATTT TCATCATAAG AAAGCACTGT ATTGATGACT	540
TGGACTTCAT TACGGAAACC TTCCGCAAAC ATAGGACGAA TCGGACGGTC AATCAAACGC	600
GCTGTCAAGG TCGCATCTGT TGAAGGACGT CCTTCACGTT TCATAAAGCC ACCAGGAAAC	660
TTCCAGCCG CATACATTTT TTCTTCGTAG TTGACTTGA GTGGGAAGAA ATCCTCAGTT	720
GCCATTTTCT GGGGATCC	738

(2) INFORMATION FOR SEQ ID NO: 384:

1394

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 657 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

CCCCCTATTT ACCGTGGACT AAAGTTGTAC AAGAAAAGTG CAAATAAGAA ATCTCCAGAT	60
TAGGAACTAT ATATGAGTTC TCTAGTCTGG AGATTTTTC AATAGACTTCG TTATTGGGCG	120
GTTACTTTTCG AAACTTTGAA AACTTCAAAA AACGGATTTT TATCGCTTTC AAATCTTTT	180
GGGGTCAAAC TCAGTAACTT ATTCGCCTTG TAGACTTCAT GACGCTCAGG GTATACTTTC	240
AAGGTCCCAA ATAGCCAAGA ATCGTCAGCG ATATTATCTG AATCATCTCC TTCTTGTCT	300
CCTTTAGTTC GCCTGAGGAC AGCCTTGACA CGCGCCAGAA TTCTCTAGGG CTAAAAGGCT	360
TGGTCAGGTA GTCATCAGCC CCTAATTCCA AGGCCAAAAC CTTATCAAAT TCATCACTTT	420
TCGCAGAAAC CATCATAATT GGAGTTTGA CGCCTTGGC TCTCAGCCGC TTACAAACTT	480
CCATGCCATC TAATTGTGGT AACATGATAT CAAGCAAGAT AAAATCAAAG GGTCTCTTTT	540
CTGCCAAAGC TAAGGCCTTC CGTCCATTG TCACCAATTG AGTAGAAAAG CCTTCCTTAC	600
TTAAATGGTA GTCAAGCAAT TTCAGAATGT GTTCTTCATC ATCCACTAAT AAGACTT	657

(2) INFORMATION FOR SEQ ID NO: 385:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 586 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

CCGCATCAGC ATCAACGAGT GCATCGGCTT CACGTCAACC AGTGCATCAG TCTCAGCAAG	60
CACCAGTGCG TCGGCTTCAG CATCAACGAG TGCCTCAGCC TCAGCAAGTA TCTCAGCGTC	120
TGAATCGGCA TCAACGAGTG CGTCAGCTCA GCAAGTACTA GTGCATCGGC TTCAGCAAGC	180
ACCAGTGCCT CGGCTTCAGC ATCAACGAGT GCCTCAGCCT CAGCAAGTAT CTCAGCGTCT	240
GAATCGGCAT CAACGAGTGC GTCACCTCAG CAAGTACTAG TGCATCAGCA TCAGCATCAA	300
CGAGTGCAAT GGCTTCAGCA AGTACCAGCG CCTCAGCTTC AGCAAGCACC AGTGCCTCAC	360
CTCAGCAAGT ACCAGCGCCT CAGCCTCAGC AAGCACCAGT GCCTCAGCTT CAGCAAGTAC	420
CAGTGCGTCA CCTCAGCATC GACAAGTGCG TCGGCTTCAG CAAGTACCTC AGCGTCTGAA	480

1395

TCAGCATCAA CGAGTGGCTC AGCTTCAGCA TCAACCAGTG CCTCAGCCTC AGCAAGTATC 540  
 AGTGGCTCAG CTTTCAGCATC AACGAGTGCG TCAGCTGCAG CAAGTA 586

## (2) INFORMATION FOR SEQ ID NO: 386:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 451 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 386:

CGTCGGCTTC AGCATCAACG AGTGCATCAG CTTTCAGCATC AACAAAGTGCT TCAGCTTCAG 60  
 CAAGTACCAG TGGCTCGGCT TCAGCATCAA CGAGTGCTTC AGTCTCAGCG TCAACCAGTG 120  
 CCTCTGAATC CGCATCAACA AGTGCCTCGG CTTTCAGCAAG CACCAGTGCT TCGGCTTCAG 180  
 CGTCAACGAG TGGCTCTGAG TCAGCATCAA CGAGTGGCTC ACCTCAGCAA GCACATCAGC 240  
 TTCTGAATCT GCATCAACCA GTGCGTCAGC TTCCGCATCA ACAAGCGCCT CGGCCTCAGC 300  
 AAGTACAAGT GCTTCAGCCT CAGCATCAAC CAGTGCATCA GCTTCAGCCT CAACAAGTGC 360  
 TTCAGCCTCA GCGTCAACCA GTGCCTCGGC TTCAGCAAGT ACCAGTGCGT CAGTTCAGCA 420  
 AGCACAAGTG CGTCAATTTA GCATCAACCA G 451

## (2) INFORMATION FOR SEQ ID NO: 387:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 425 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 387:

TCTCAGCAAG CACCATTGCG TCGGCTTCAT CAAGCACCAG CGCGTTTGAA TCCGCATCAA 60  
 CCAGTGCTTC AGCTTCAGCC AAGTTACCTC AGCATCTGAA TCAGCATCAA CAAGTGCATC 120  
 GGCTTCAGCA AGCACAAGTG CTTTCAGCTCA GCAAGTATCT CAGCGTCTGA ATCGGCATCA 180  
 ACGAGTGCGT CCGCTTCAGC AAGTACTAGC GCCTCAGCAT CAGCGTCAAC AAGTGCTTCG 240  
 GCTTCAGCGT CAACGAGTGC GTCTGAGTCA GCATCAACGA GTACGTCAGC CTCAGCAAGC 300  
 ACATCAGCTT CTGAATCTGC ATCAACCAGT GCGTCAGCCT CAGCATCGAC AAGCGCCTCA 360  
 GCTTCAGCAA GTACCAGTGC GTCAGCCTCA GCAAGTACCA GTGCTTCAGC CTCAGCGTCG 420

1396

ACAAG

425

## (2) INFORMATION FOR SEQ ID NO: 388:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 572 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

AGAGGATCCC CGGATCCTCA GTCGCTGAGA TAACTCCTTT GGGCTTGTTT ATCATGTAGT	60
AGACAAATC TTCATACTCC AACACTTGCC CATTTTATGC GAATCTCATC TATTTTCTCT	120
TTTTTTTGCA ATTTAGCTGA TTTTCTTTT TTACCATTTA CAGTCACGCG CCCAGCCTTG	180
AGCAAGTTT TGACCTCAGT CCGACTTCCC ACCGCACAGG CAACTAAAAA TTTATCTAAT	240
CTCATAGAAC TATTATATCA TATCAAAAGG AGGCTAGTAC AATGACCAAC CTCCTTTTCG	300
TTTCATACTC TTCAAAAATC TCTTCAAACC GCGTCAACGT CGCCTTGCCG TATATATGTT	360
ACTGACTTCG TCAGTTCTAT CTGCAACCTC AAAACAGTGT TTTGAGCTGA CTTCGTCAGT	420
TCTATCTGCA ACCTCAAAGC AGTGCTTTGA GCATCCTGCG GCTAGTTTCC KAGTKTGCTC	480
TTTGATTTC ATTGAGTATC AGATTAGGA AATTAAGTTC CTCGKCTCCA AAAAAGAGCT	540
AAAACAATCA AGGCTCCTAA AATCGCTGGG AT	572

## (2) INFORMATION FOR SEQ ID NO: 389:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 505 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

CAACAAGTGC CTCGGCTTCA GCATGCACAA GTGCTTCAGC TTCAGCATGT ACCTGAGCGT	60
CTGAATCAGC ATCAACGTGT GCGTCCGCTT CAGCATGTAC TGCTGCCTCA GCATCAGCGT	120
CAACAAGTGC TTCGGCTTCA GCGTCAACGA GTGCGTCTGA GTCAGCATCA ACGAGTACGT	180
CAGCCTCAGC AAGCACATCA GCTTCTGAAT CTGCATCAAC CAGTGCCTCA GCCTCAGCAT	240
CGACAAGCGC CTCAGCTTCA GCAAGTACCA GTGCGTCAGC CTCAGCAAGT ACCAGTGCTT	300
CAGCCTCAGC GTCGACAAGT GCGTCGGCCT CAACCAAGTGC ATCTGAATCG GCATCAACCA	360
GTGCGTCAGC CTCAGCAAGT ACTAGCGCCT CAGCCTCAGC ATCAACGAGT GCGTCCGCTT	420

1397

CAGCAAGTAC TAGTGCATCA GCATCAGCAT CAACGAGTGC ATCGGCTTCA GCAAGTACCA 480  
GCGCCTCAGC TTCAGCAAGC ACCGG 505

(2) INFORMATION FOR SEQ ID NO: 390:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 447 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

GCTAAGACTA CCTCATTAGG GGCATAGGCT GCTAAAATAA CTGCAGCTGT GGTTAATGAC 60  
AATACTGTAC TTTTTCAT TTTAATTCCT TACATATTTA TATAACTTCC AATAGATAAT 120  
AAACTTTAAC TTGCTAGCC TTGTTATAA AAAGTTTAC TAAGTATTAT CTAGGAAATA 180  
GAGTAGTACA TTTATATATA ATTGTTATCT CTCTATAAAA ACAGTATATC ATTTAAAAAA 240  
ATTTAAGTCA AAAAAATTAA CATTAGTTAA TTTATTTTTT AGCACACATT AAAAAATAAG 300  
ATTAGTACTC AATGAAAATC AAAGAGCAAA CTAGGAAACT AGCCGCAGAT TGCTCAAAAC 360  
AGTGTTTGA GGTGTAGAT GGAATGACGT AGTCAGCTCA AAACACTGTT TTGAAGTTGT 420  
GGATAGAACT GACGAAGTCG GTACCGA 447

(2) INFORMATION FOR SEQ ID NO: 391:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 572 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 391:

AGCACTTGTC GTTGAATTCT ACAACAAAAT GTTGTAAATAT TTTATTGAAT AAGTAGGCC 60  
TTGATATTAA GCACTTTGGG ACGTCTCCC TTAGTGCTTT TTTGATTCT CTTAGTATCC 120  
AGCTATAATC GTTGAGACAT AACTAGACCG ATATAGTCCA AAGTGATATA GTAAATGAA 180  
CCAAAAATAG TACACAATGT GGTATAATCC TTTATGGCA TATTCAATAG ATTTTCGTAA 240  
AAAAGTTCTC TCTTATTGTG AGCGAACAGG TAGTATAACA GAAGCATCAC ACGTTTCCA 300  
AATCTCACGT AATACCATTT ATGGCTGGTT AAAGCTAAAA GAGAAAACAG GAGAGCTAAA 360  
CCACCAAGTA TAGTGATTG AATCTATAAC AGTACACCTT GGCTGCTAAA ATATTTCTAT 420

1398

AAATTAATTT GACTTTCCTG ATAGAGATGT TCACATCTTA TTTCAAATA CTATATAAGT	480
TCTATAATCT CTTTATAAGA TTTGCCCATC AGACAAAATA GAACGATTG AAGGCGTTTA	540
TGATATTTAG CTGTACGAGA GTCTTTTAAA AG	572

---

MISSING UPON TIME OF PUBLICATION

**DENMARK**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Danish Patent Office), or has been finally decided upon by the Danish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Danish Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Danish Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Danish Patent Office or any person approved by the applicant in the individual case.

**SWEDEN**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Swedish Patent Office), or has been finally decided upon by the Swedish Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the International Bureau before the expiration of 16 months from the priority date (preferably on the Form PUT/RO/134 reproduced in annex Z of Volume I of the PCT Applicant's Guide). If such a request has been filed by the applicant, any request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Swedish Patent Office or any person approved by the applicant in the individual case.

**UNITED KINGDOM**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for the International publication of the application.

**NETHERLANDS**

The applicant hereby requests that until the date of a grant of a Netherlands patent or until the date on which the application is refused or withdrawn or lapse, the microorganism shall be made available as provided in Rule 31F(1) of the Patent Rules only by the issue of a sample to an expert. The request to this effect must be furnished by the applicant with the Netherlands Industrial Property Office before the date on which the application is made available to the public under Section 22C or Section 25 of the Patents Act of the Kingdom of the Netherlands, whichever two dates occurs earlier.



**SINGAPORE**

The applicant hereby requests that the furnishing of a sample of a microorganism shall only be made available to an expert. The request to this effect must be filed by the applicant with the International Bureau before the completion of the technical preparations for international publication of the application.

**NORWAY**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Norwegian Patent Office), or has been finally decided upon by the Norwegian Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art. The request to this effect shall be filed by the applicant with the Norwegian Patent Office not later than at the time when the application is made available to the public under Sections 22 and 33(3) of the Norwegians Patents Act. If such a request has been filed by the applicant, any request made by a third party for the furnishing of a sample shall indicate the expert to be used. That expert may be any person entered on a list of recognized experts drawn up by the Norwegian Patent Office or any person approved by the applicant in the individual case.

**AUSTRALIA**

The applicant hereby gives notice that the furnishing of a sample of a microorganism shall only be effected prior to the grant of a patent, or prior to the lapsing, refusal or withdrawal of the application, to a person who is a skilled addressee without an interest in the invention (Regulation 3.25(3) of the Australian Patents Regulations).

**FINLAND**

The applicant hereby requests that, until the application has been laid open to public inspection (by the National Board of Patents and Registration), or has been finally decided upon by the National Board of Patents and Registration without having been laid open to public inspection, the furnishing of a sample shall only be effected to an expert in the art.

**ICELAND**

The applicant hereby requests that, until the application has been laid open to public inspection (by the Icelandic Patent Office), or has been finally decided upon by the Icelandic Patent Office without having been laid open to public inspection, the furnishing of a sample shall only be effected in the art.

**What Is Claimed Is:**

- 25           1. Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID NOS:1-391, a representative fragment thereof or a nucleotide sequence at least 95% identical to a nucleotide sequence depicted in SEQ ID NOS:1-391.
- 30           2. Computer readable medium having recorded thereon any one of the fragments of SEQ ID NOS:1-391 depicted in Tables 2 and 3 or a degenerate variant thereof.
- 35           3. The computer readable medium of claim 1, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
- 40           4. The computer readable medium of claim 3, wherein said medium is selected from the group consisting of a floppy disc, a hard disc, random access memory (RAM), read only memory (ROM), and CD-ROM.
5. A computer-based system for identifying fragments of the *Streptococcus pneumoniae* genome of commercial importance comprising the following elements:
- 45           a) a data storage means comprising the nucleotide sequence of SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-391;
- b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and
- c) retrieval means for obtaining said homologous sequence(s) of step (b).
- 50           6. A method for identifying commercially important nucleic acid fragments of the *Streptococcus pneumoniae* genome comprising the step of comparing a database comprising the nucleotide sequences depicted in SEQ ID NOS:1-391, a representative fragment thereof, or a nucleotide sequence at least 95% identical to a nucleotide sequence of SEQ ID NOS:1-391 with a target sequence to obtain a
- 55           nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected.

60 7. A method for identifying an expression modulating fragment of  
*Streptococcus pneumoniae* genome comprising the step of comparing a database  
comprising the nucleotide sequences depicted in SEQ ID NOS:1-391, a  
representative fragment thereof, or a nucleotide sequence at least 95% identical to  
the nucleotide sequence of SEQ ID NOS:1-391 with a target sequence to obtain a  
nucleic acid molecule comprised of a complementary nucleotide sequence to said  
65 target sequence, wherein said target sequence comprises sequences known to  
regulate gene expression.

8. An isolated protein-encoding nucleic acid fragment of the *Streptococcus*  
*pneumoniae* genome, wherein said fragment consists of the nucleotide sequence of  
70 any one of the fragments of SEQ ID NOS:1-391 depicted in Tables 2 and 3, or a  
degenerate variant thereof.

9. A vector comprising any one of the fragments of the *Streptococcus*  
*pneumoniae* genome SEQ ID NOS:1-391 depicted in Tables 2 and 3 or a  
75 degenerate variant thereof.

10. An isolated fragment of the *Streptococcus pneumoniae* genome,  
wherein said fragment modulates the expression of an operably linked open reading  
frame, wherein said fragment consists of the nucleotide sequence from about 10 to  
80 200 bases in length which is 5' to any one of the open reading frames depicted in  
Tables 2 and 3 or a degenerate variant thereof.

11. A vector comprising any one of the fragments of the *Streptococcus*  
*pneumoniae* genome of claim 8.  
85

12. An organism which has been altered to contain any one of the  
fragments of the *Streptococcus pneumoniae* genome of claim 8.

13. An organism which has been altered to contain any one of the  
90 fragments of the *Streptococcus pneumoniae* genome of claim 10.

14. A method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 10 to 100 bases 5' to any one of the fragments of the *Streptococcus pneumoniae* genome depicted in SEQ ID NOS:1-391 and Tables 2 and 3 or a degenerate variant thereof.

15. An isolated nucleic acid molecule encoding a homolog of any of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and Tables 2 and 3, wherein said nucleic acid molecule is produced by a process comprising steps of:

a) screening a genomic DNA library using as a probe a target sequence defined by any of SEQ ID NOS:1-391 and Tables 2 and 3, including fragments thereof;

b) identifying members of said library which contain sequences that hybridize to said target sequence; and

c) isolating the nucleic acid molecules from said members identified in step (b).

16. An isolated DNA molecule encoding a homolog of any one of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and Tables 2 and 3, wherein said nucleic acid molecule is produced a process comprising steps of:

a) isolating mRNA, DNA, or cDNA produced from an organism;

b) amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from said fragment of said *Streptococcus pneumoniae* genome to prime said amplification;

c) isolating said amplified sequences produced in step (b).

17. An isolated polypeptide encoded by any of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and depicted in Table 2 and 3 or by a degenerate variant of said fragments.

18. An isolated polynucleotide molecule encoding any one of the polypeptides of claim 17.

19. An antibody which selectively binds to any one of the polypeptides of claim 17.

130

20. A method for producing a polypeptide in a host cell comprising the steps of:

135

a) incubating a host containing a heterologous nucleic acid molecule whose nucleotide sequence consists of any one of the fragments of the *Streptococcus pneumoniae* genome of SEQ ID NOS:1-391 and depicted in Tables 2 and 3, under conditions where said heterologous nucleic acid molecule is expressed to produce said protein, and

b) isolating said protein.

Figure 1

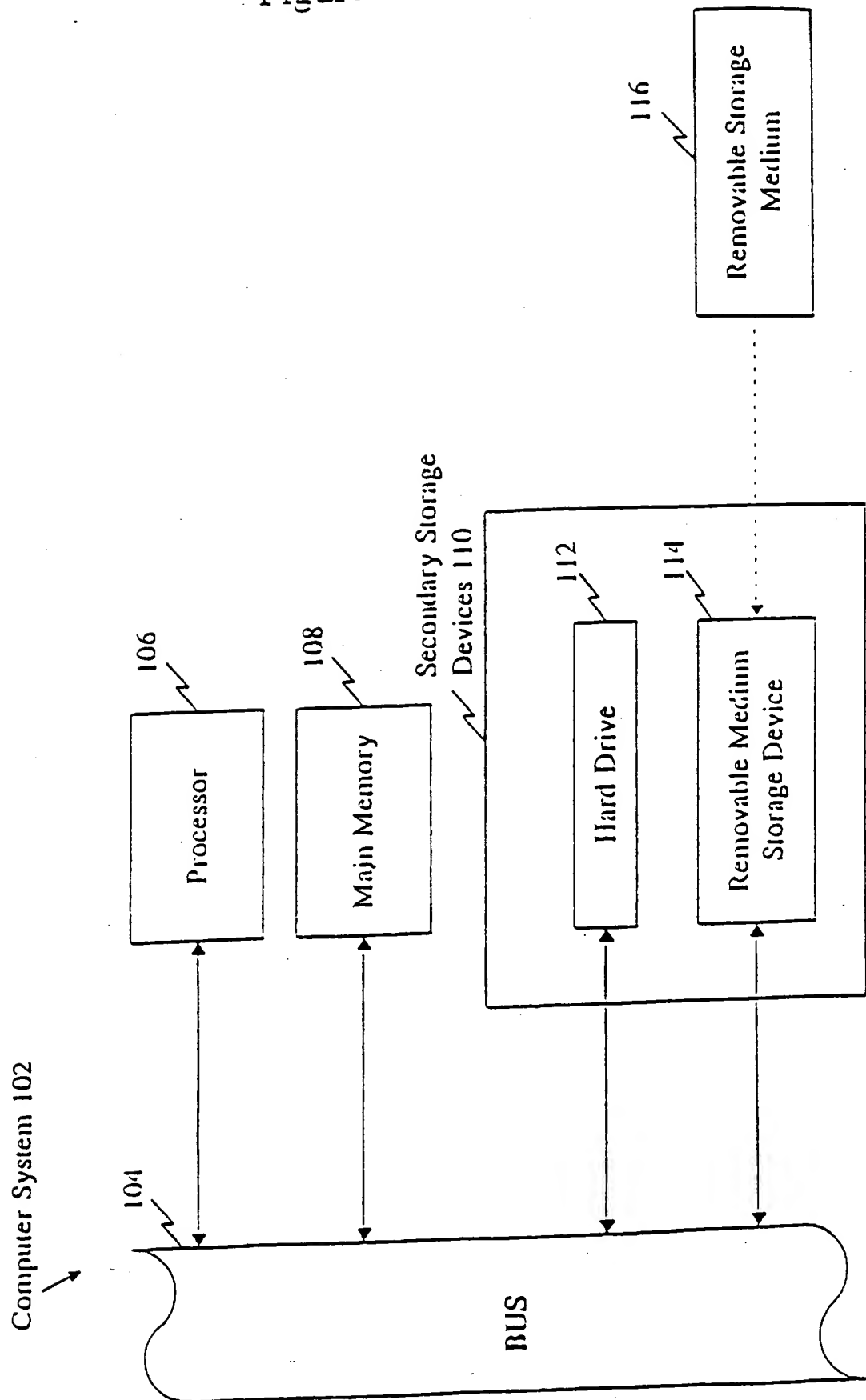
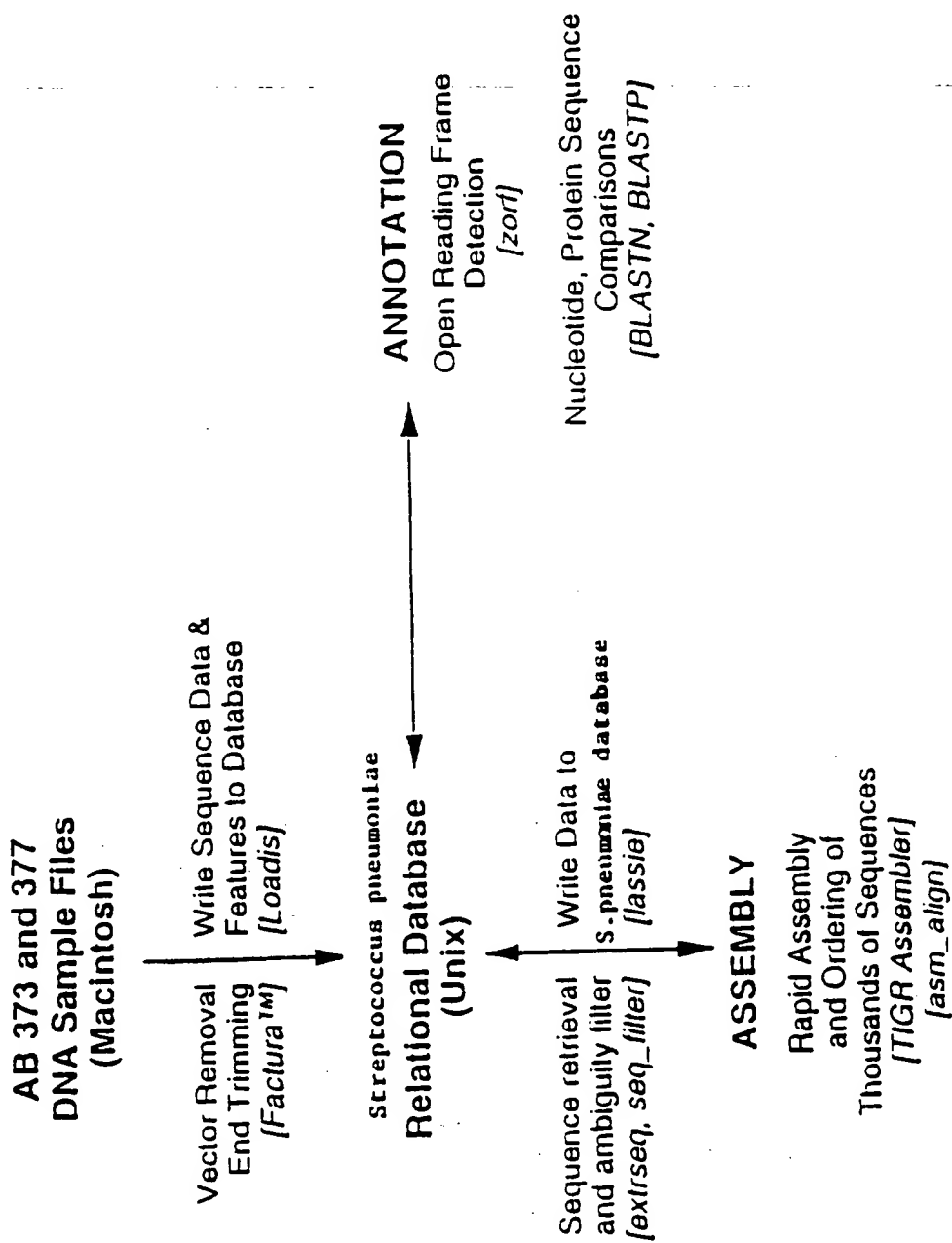


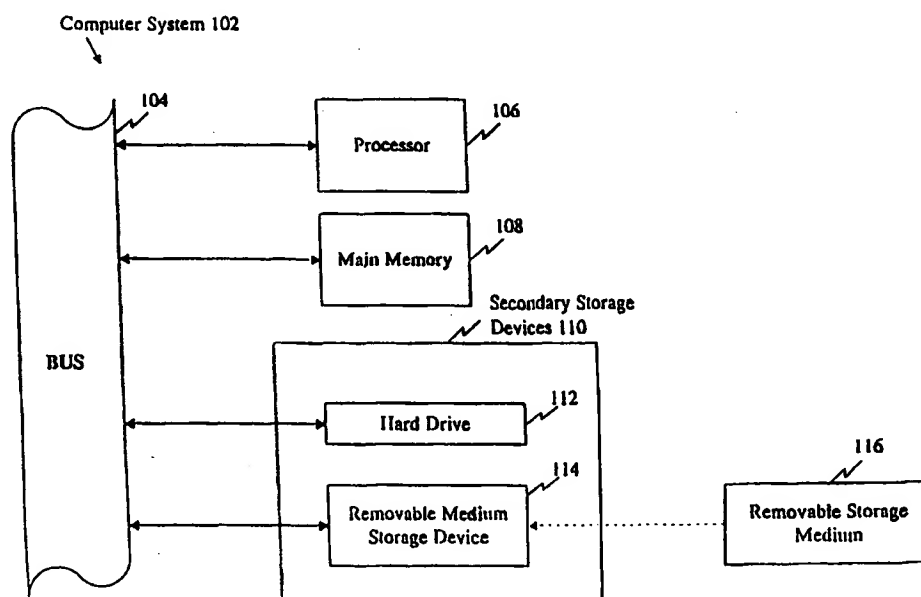
Figure 2





## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 15/31, C07K 14/315, 16/12, C12Q 1/68</b>		(11) International Publication Number: <b>WO 98/18931</b>
A3		(43) International Publication Date: 7 May 1998 (07.05.98)
(21) International Application Number: PCT/US97/19588		(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).
(22) International Filing Date: 30 October 1997 (30.10.97)		
(30) Priority Data: 60/029,960 31 October 1996 (31.10.96) US		
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(72) Inventors; and		
(75) Inventors/Applicants (for US only): KUNSCH, Charles, A. [US/US]; 2398B Dunwoody Crossing, Atlanta, GA 30338 (US). CHOI, Gil, H. [KR/US]; 11429 Potomac Oaks Drive, Rockville, MD 20850 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe Court, Carlsbad, CA 92009 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). BARASH, Steven, C. [US/US]; 582 College Parkway #303, Rockville, MD 20850 (US). FANNON, Michael [US/US]; 13501 Rippling Brook Drive, Silver Spring, MD 20850 (US). DOUGHERTY, Brian, A. [US/US]; 708 Meadow Field Court, Mount Airy, MD 21771 (US).		
		Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
		(88) Date of publication of the international search report: 20 August 1998 (20.08.98)

(54) Title: *STREPTOCOCCUS PNEUMONIAE* POLYNUCLEOTIDES AND SEQUENCES

## (57) Abstract

The present invention provides polynucleotide sequences of the genome of *Streptococcus pneumoniae*, polypeptide sequences encoded by the polynucleotide sequences, corresponding polynucleotides and polypeptides, vectors and hosts comprising the polynucleotides, and assays and other uses thereof. The present invention further provides polynucleotide and polypeptide sequence information stored on computer readable media, and computer-based systems and methods which facilitate its use.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

## INTERNATIONAL SEARCH REPORT

In. ational Application No

PCT/US 97/19588

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/31 C07K14/315 C07K16/12 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 96 33276 A (HUMAN GENOME SCIENCES INC ;UNIV JOHNS HOPKINS (US)) 24 October 1996 see claims 1-7	1-7
A	ALTSCHUL S F ET AL: "BASIL LOCAL ALIGNMENT SEARCH TOOL" JOURNAL OF MOLECULAR BIOLOGY, vol. 215, 1990, pages 403-410, XP000604562 cited in the application see the whole document	1-7



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*Z\* document member of the same patent family

Date of the actual completion of the international search

27 March 1998

Date of mailing of the international search report

0 8. 07. 98

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

HORNIG H.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/19588

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	W.R. PEARSON AND D.J. LIPMAN: "Improved tools for biological sequence comparison" PROC. NATL. ACAD. SCI., vol. 85, April 1988, NATL. ACAD. SCI., WASHINGTON, DC, US;, pages 2444-2448, XP002060460 cited in the application see the whole document ---	1-7
A	WO 95 06732 A (UNIV ROCKEFELLER ;MASURE H ROBERT (US); PEARCE BARBARA J (US); TUO) 9 March 1995 see the whole document ---	1-7
A	WO 95 31548 A (UAB RESEARCH FOUNDATION ;YOTHER JANET (US); DILLARD JOSEPH P (US)) 23 November 1995 see the whole document ---	1-7
A	WO 95 14712 A (RES CORP TECHNOLOGIES INC) 1 June 1995 see the whole document ---	1-7
A	WO 96 05859 A (AMERICAN CYANAMID CO) 29 February 1996 see the whole document ---	1-7
A	WO 93 10238 A (US HEALTH) 27 May 1993 see the whole document ---	1-7
A	EP 0 687 688 A (UNIV OVIEDO ;UNIV LEICESTER (GB)) 20 December 1995 see the whole document ---	1-7
A	EP 0 622 081 A (UAB RESEARCH FOUNDATION) 2 November 1994 see the whole document -----	1-7

# INTERNATIONAL SEARCH REPORT

...ernational application No.

PCT/US 97/ 19588

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although claims 1-4 could be, at least partially be considered as a mere presentation of information Rule 39.1(v), and claims 5-7 at least partially as a computer program, Rule 39.1(vi)PCT, the search has been carried out as far as possible in our systematic documentation.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see continuation-sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-7

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

## 1. Claims: 1-7

Computer readable medium having recorded thereon the nucleotide sequence depicted in SEQ ID nos. 1-391, a representative fragment thereof or a nucleotide sequence at least 95% identical to a nucleotide sequence depicted in SEQ ID nos. 1-391; a computer-based system for identifying fragments of the *Streptococcus pneumoniae* genome of commercial importance comprising: a) a data storage means comprising said nucleotide sequence(s); b) search means for comparing a target sequence to the nucleotide sequence of the data storage means of step (a) to identify homologous sequence(s), and c) retrieval means for obtaining said homologous sequence(s) of step (b); a method for identifying commercially important nucleic acid fragments of the *Streptococcus pneumoniae* genome comprising the step of comparing a database comprising said nucleotide sequence(s) with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence is not randomly selected; a method for identifying an expression modulating fragments of the *Streptococcus pneumoniae* genome comprising the step of comparing a database comprising said nucleotide sequence(s) with a target sequence to obtain a nucleic acid molecule comprised of a complementary nucleotide sequence to said target sequence, wherein said target sequence comprises sequences known to regulate gene expression;

## 2. Claims: (8-20) partially

An isolated protein-encoded nucleic acid fragment of the *Streptococcus pneumoniae* genome, wherein said fragment consists of the nucleotide sequence of the fragment of SEQ ID no.1 depicted in Tables 2 and 3, or a degenerate variant thereof; a vector comprising the fragment of the *Streptococcus pneumoniae* genome SEQ ID no.1; an isolated fragment of the *Streptococcus pneumoniae* genome, wherein said fragment modulates the expression of an operably linked open reading frame, wherein said fragment consists of the nucleotide sequence from about 10 to 200 bases in length which is 5' to any one of the open reading frame of SEQ ID no.1 depicted in Tables 2 and 3 or a degenerate variant thereof; a method for regulating the expression of a nucleic acid molecule comprising the step of covalently attaching to said nucleic acid molecule a nucleic acid molecule consisting of the nucleotide sequence from about 10 to 100 bases 5' to any one of the open reading frame of SEQ ID no.1 and Tables 2 and 3 or a degenerate variant thereof; an isolated nucleic acid molecule encoding a homolog of SEQ ID no.1; an isolated polypeptide encoded by SEQ ID no.1 and depicted in Table 2 and 3; an antibody which selectively binds to any one of said polypeptides, a method for producing a polypeptide in a host cell comprising a) incubating a host containing a heterologous nucleic acid

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

molecule whose nucleotide sequence consists of SEQ ID no.1 and depicted in Table 2 and 3, under conditions where said heterologous nucleic acid molecule is expressed to produce said protein, and b) isolating said protein;

3-392. Claims:(8-20) partially

Idem as subject 2 but limited to each of the sequences of SEQ ID no. 2 to 391;

For the sake of conciseness, the second subject matter is explicitly defined, the other subject matters are defined by analogy hereto.

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/US 97/19588

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9633276 A	24-10-96	AU 5552396 A EP 0821737 A	07-11-96 04-02-98
WO 9506732 A	09-03-95	AU 7680994 A CA 2170726 A EP 0721506 A FI 960977 A JP 9504686 T NO 960839 A	22-03-95 09-03-95 17-07-96 30-04-96 13-05-97 19-04-96
WO 9531548 A	23-11-95	AU 2638595 A EP 0804582 A	05-12-95 05-11-97
WO 9514712 A	01-06-95	US 5474905 A	12-12-95
WO 9605859 A	29-02-96	US 5565204 A AU 3363695 A CA 2198251 A EP 0778781 A JP 10504717 T	15-10-96 14-03-96 29-02-96 18-06-97 12-05-98
WO 9310238 A	27-05-93	AU 3065892 A	15-06-93
EP 0687688 A	20-12-95	ES 2075803 A ES 2088820 A WO 9516711 A	01-10-95 16-09-96 22-06-95
EP 0622081 A	02-11-94	AU 682018 B AU 5769694 A CA 2116261 A FI 941695 A JP 7126291 A NO 941420 A US 5679768 A ZA 9401584 A	18-09-97 27-10-94 21-10-94 21-10-94 16-05-95 21-10-94 21-10-97 12-10-94